Title
Third International Conference/Workshop on Integrating GIS and Environmental Modeling

Author
National Center for Geographic Information and Analysis

Publication Date
1996

NCGIA National Center for Geographic Information and Analysis

Third International Conference/Workshop on Integrating GIS and Environmental Modeling

MAIN MENU

NCGIA National Center for Geographic Information and Analysis

Third International Conference/Workshop on Integrating GIS and Environmental Modeling

Santa Fe, New Mexico, U.S.A.

January 21-25, 1996

BACKGROUND

The Third International Conference/Workshop on Integrating GIS and Environmental Modeling was held under the auspices of the U.S. National Center for Geographic Information and Analysis in Santa Fe, NM from Sunday January 21 to Thursday January 25, 1996. The conference followed previous meetings in Boulder, Colorado in 1991 and Breckenridge, Colorado in 1993, each attended by over 600 participants.

The conference had three interrelated objectives:

1. To review the current status of digital geographic information for environmental modeling, with particular emphasis on the technical and institutional issues affecting its usefulness and accessibility.

2. To review progress, with emphasis on the period since the previous conference, in the development of environmental models, and in the exploitation of geographic information technologies, particularly GIS, to support modeling.

3. To identify areas where progress in the integration of GIS and environmental modeling is likely to be made in the next few years, through improvements in technology, institutional structures, and modeling methods.

The program was organized around three themes, each addressing one of the three objectives:

Data Issues: topics include but are not limited to data quality, reports on new data sources, spatial data infrastructures, new technologies for data access including digital spatial data libraries, intellectual property issues, economics of spatial data provision, metadata and format standards, methods of discretization, data modeling and data structures, methods of spatial analysis including interpolation and regionalization, integration of GIS and remote sensing.

Progress in Modeling: reports and demonstrations of progress in integrating GIS and environmental modeling in such fields as atmospheric science, ecology, oceanography, hydrology, spatial decision support, biodiversity, water and air quality, risk assessment, global environmental change, coupled systems or integrated modeling, and appropriate contributions to the global modeling of carbon, trace gas fluxes, etc.

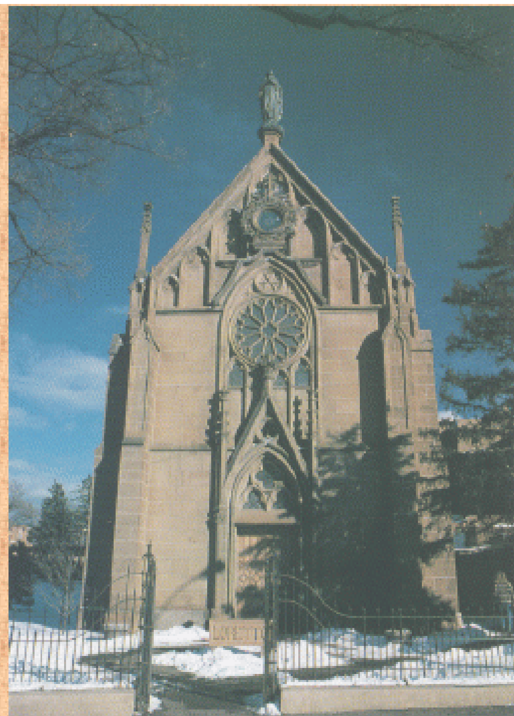
New Research Frontiers: discussions or demonstrations of research offering potential for new approaches to environmental modeling with GIS, including such topics as cellular automata models, modeling languages, computational modeling systems, new approaches to data modeling including time, 3D, 4D, and global modeling, object oriented systems, and agent- or event-based programming.

In addition to sessions on each of these themes, the conference followed the pattern of previous conferences by including workshops and tutorials on significant topics, informal discussion sessions, poster sessions and demonstrations.

ORGANIZING COMMITTEE:

Michael Goodchild, NCGIA
Louis Steyaert, USGS
Bradley Parks, University of Colorado
Michael Crane, USGS
Carol Johnston, University of Minnesota
John Wilson, Montana State University
Denice Shaw, US EPA
Sandi Glendinning, NCGIA

CONFERENCE STUDENT SCHOLARSHIPS



The NCGIA Conference Secretariat
National Center for Geographic Information and Analysis Phelps Hall 3510
University of California
Santa Barbara, CA 93106-4060, USA

Go Back to the Main Menu

NCGIA National Center for Geographic Information and Analysis

Third International Conference/Workshop on Integrating GIS and Environmental Modeling

January 21-25, 1996
Santa Fe, New Mexico, USA

PROGRAM

SUNDAY - January 21, 1996

Workshops

- Terrain Analysis Methods for Environmental Characterization and Modeling, John P. Wilson, Montana State University<184>
- Techniques for the Organization and Management of Network-Accessible Repositories of Datasets, Data Descriptions, and Data Transformation, Ray Ford and M. Sweet, University of Montana<167>
- The Development and Application of a GIS-Based Prehistoric Resource Distribution Model in Archaeological Research, Karl Benedict, Louis Scuderi, University of New Mexico, Albuquerque, and Richard Watson, San Juan College, Farmington
- Land Use/Land Transformation Modeling Workshop, Bryan C. Pijanowski, Michigan State University; Brad Parks, Cooperative Insitute for Research in Environmental Sciences (CIRES). Presenters:
    - Communicating Scientific Findings to the General Public, Stacy Hoppen, Keith C. Clarke, Leonard C. Clarke, Leonard J. Gaydos, and William Acevedo.
- The Swarm Modelling System, Chris Langton, Santa Fe Institute.
- New Research Frontiers in 3D Subsurface Characterization, Frank A. D'Agnese, Claudia C. Faunt, US Geological Survey; and A. Keith Turner, Colorado School of Mines<173>
- GIS, Knowledge Bases, Groupware, and Decision Support: Linked Technologies for Resource Management, Brenda Faber, John Calkins, Keith Reynolds
- GRASS Special Interest Group
- On the Use of GIS in Mapping Spatio-Temporal Patterns of Environmental Variability, Assaf Anyamba, Clark University
- Multi-Criteria Modelling in GIS using Fuzzy Measures, J. Ronald Eastman and Hong Jiang, Clark University
- Design Issues for GIS-Based Environmental Modelling: Socio-Economic and Biophysical Perspectives, Richard Aspinall, MLURI, Aberdeen, and Brian Lees, Australian National University, Canberra<6>

Evening

- Welcome Address, John E. Estes, University of California, Santa Barbara
- Keynote, Darrel L. Williams, NASA

MONDAY - January 22, 1996

Plenaries

1. GIS-Based Land Surface/Subsurface Modeling: New Potential for New Models? John Wilson, Montana State University
2. Internet Access to Spatial Data: Present and Future of the National Geospatial Data Clearinghouse, Doug Nebert, US Geological Survey

CLIMATE/WEATHER INTERPOLATION - SESSION 1

Chair, John Wilson, Montana State University

1. Comparison of Anusplin, MT-CLIM-3D and PRISM preciptation estimates, John Wilson and S.T. Stillman, Montana State University
2. Spatial Modelling of Climatic Variables on a Continental Scale, Jennifer Kesteven, Australian National University; and Michael Hutchinson, Australian National University<108>
3. A Comparison of Spatial Interpolation Techniques in Temperature Estimation, Fred Collins and P.V. Bolstad, IBM
4. Smart Interpolation of Climate Variables, Cort Willmott, University of Delaware (abstract only)

INTERNET/DATA ACCESS 1 - SESSION 2

Chair, Barbara Buttenfield, University of Colorado

1. Tools for Browsing Environmental Data: The Alexandria Digital Library Interface, Barbara P. Buttenfield, University of Colorado; and Mark P. Kumler, University of Colorado<23>
2. Environmental Data Access in New Zealand 1985-1995: user-pays to open-access and the WWW, James Barringer, Landcare Research New Zealand Ltd.; Julian Cone, Landcare Research New Zealand Ltd.; Robert Gibb, Landcare Research New Zealand Ltd.; Hamish Heke, Landcare Research New Zealand Ltd.; David Medyckyj-Scott, Landcare Research New Zealand Ltd.; Peter Newsome, Landcare Research New

Zealand Ltd.; and Janice Willoughby, Landcare Research New Zealand Ltd.<11>

3. A Structure for Organizing Metadata Collection, Kate Beard, University of Maine
4. UDK: A European Environmental Data Catalogue, Oliver Guenther, Institut fur Wirtschaftsinformatik; Helmut Lessing, Nieders Swobodachsisches Umweltministerium; Walter Swoboda, Nieders Swobodachsisches Umweltministerium; presented by R. Nikolai<74>

LAND SURFACE CHARACTERIZATION: VEGETATION 1 - SESSION 3

Chair, Lou Steyaert, U.S. Geological Survey

1. The U.S. Geological Survey's Land Cover Characterization Program, Nick Van Driel, EROS Data Center; and Tom Loveland, EROS Data Center
2. Land Use History of North America -- Need for a Continential Synthesis, Thomas D. Sisk, National Biological Service
3. Constructing Detailed Vegetation Databases from Field Data and Airborne Videography, Carol A. Johnston, University of Minnesota; Carol A. Sersland, University of Minnesota; John Bonde, University of Minnesota; Deb Pomroy-Petry, University of Minnesota; and Paul Meysembourg, University of Minnesota
4. GIS, Spatial Statistical Graphics, and Forest Health, James J. Majure, Iowa State University; Noel Cressie, Dianne Cook, and Jurgen Symanzik

INTERNATIONAL AGRICULTURAL DEVELOPMENT - SESSION 4

Chair, John Corbett, International Centre for Research in Agroforestry

1. The Changing Face of Agroecosystem Characterization: Models and Spatial Data, the Basis for Robust Agroecosystem Characterization, John D. Corbett, International Centre for Research in AgroForestry<37>
2. The Development of a Topographic and Climate Database for Africa, M.F. Hutchinson, Australian National University; H.A. Nix, Australian National University; J.P. McMahon, Australian National University; K.D. Ord, Australian National University<78>
3. Predicting Plant Growth: Where will it grow? How well will it grow?, Trevor H. Booth, CSIRO Division of Forestry<37>
4. Deforestation in Two Brazilian Amazon Colonies: Analysis Combining Farmer Interviews and GIS, Sam Fujisaka, Centro Internacional de Agricultura Tropical (CIAT), Columbia; Nick Thomas, Centro Internacional de Agricultura Tropical (CIAT), Columbia; and Euan Crawford, Centro Internacional de Agricultura Tropical (CIAT), Columbia<37>

REMOTE SENSING AND IMAGE PROCESSING - SESSION 5

Chair, Keith Clarke, Hunter College, City University of New York

1. Design and Documentation of a Baltimore-Washington Regional Spatial Database Testbed for Environmental Model Calibration and Verification, Timothy W. Foresman, University of Maryland Baltimore County; Helen V. Wiggins, University of Maryland Baltimore County; Dana L. Porter, University of Maryland Baltimore County; Penny Masuoka, University of Maryland Baltimore County; William Acevedo, U.S. Geological Survey
2. Image Navigation for Wildland Fire Location Mapping, Loey Knapp, IBM Corporation; Patricia Andrews, USDA Forest Service; and John Turek, IBM T.J. Watson Research Lab<200>
3. Prolegomena for a Genetic Algorithm of Ecosystem Evolution, Stan Morain and Amelia Budge, University of New Mexico
4. Image Rectification with Radial Basis Functions: Application to RS/GIS Data Integration, David N. Fogel, University of California Santa Barbara

INTERNET/DATA ACCESS 2 - SESSION 6

Chair, Richard Aspinall, Macauley Land Use Research Institute

1. A Federation Architecture for an Environmental Information System incorporating GIS, the World-Wide Web, and CORBA, Arne Koschel, Forschungszentrum Informatik; Ralf Kramer, Forschungszentrum Informatik; Ralf Nikolai, Forschungszentrum Informatik; Wilhelm Hagg, Universität Karlsruhe; Joachin Wiesel, Universität Karlsruhe; Heiko Jacobe, Universität Karlsruhe<98>
2. Developing Internet-Based User Interfaces for Improving Spatial Data Access and Usability, Chun Sheng Li, University of Manchester; David Bree, University of Manchester; Adrian Moss, Manchester Metropolitan University; James Petch, Manchester Metropolitan University<105>
3. Serving GIS Data Through the World Wide Web, James Darrell McCauley, Case Corporation; Kumar C.S. Navulur, Purdue University; Bernard A. Engel, Purdue University; Raghavan Srinivasan, Blackland Research Center, Texas Agricultural Experiment Station<45>
4. A Network-Accessible Repository for the Characteriztaion of Spatial Ecosystem Components, Michael Sweet, University of Montana; Ray Ford, University of Montana; Ron Righter, University of Montana<165>

LAND SURFACE CHARACTERIZATION: SOILS - SESSION 7

Chair, Doug Miller, The Pennsylvania State University

1. Progress in Soil-landscape Modelling and Spatial Prediction of Soil Attributes for Environmental Models, Paul Gessler, CSIRO Division of Soils; Neil McKenzie, CSIRO Division of Soils; and Michael Hutchinson, Australian National University<70>
2. An Alternate Paradigm for Representing Soils Data and Data Quality Information, Bheshem Ramlal, University of Maine; and Kate Beard, University of Maine<154>
3. Using a Using a SAR image and a Decision Support System to Model Spatial Distribution of Soil Water in a GIS framework, Andrew S. Rogowski, U.S. Department of Agriculture; and Edwin T. Engman, NASA/Goddard Space Flight Center<156>
4. Towards a Methodology for Selecting a "Characteristic" Sample from an Existing Database: An Evolutionary Approach, C.H. Jarvis, N. Stuart, University of Edinburgh; J. Kelsey, and R.H.A. Baker, Ministry for Agriculture, Food & Fisheries

LAND SURFACE CHARACTERIZATION: DEMs - SESSION 8

Chair, Robert De Sawal, U.S. Geological Survey

1. A Locally Adaptive Approach to the Interpolation of Digital Elevation Models, M.F. Hutchinson, Australian National University<77>
2. Development of Continental Scale DEMs and Extraction of Hydrographic Features, Kristine L. Verdin and Susan K. Jenson, Hughes STX, EROS Data Center
3. Assessing Uncertainty in Catchment Boundary Delimitation, David R. Miller, Macauley Land Use Research Institute; and Jane G. Morrice, Macauley Land Use Research Institute<132>
4. Towards an Understanding of Landscape Scale and Structure, John C. Gallant, Australian National University; and Michael F. Hutchinson, Australian National University<68>

DATA INTEGRATION / NEW DATA SOURCES - SESSION 9

Chair, Keith Clarke, Hunter College City University of New York

1. Generic Data Exchange - Integrating Models and Data Providers, Dean Djokic, The University of New South Wales; Andrew Coates, The University of New South Wales; and James E. Ball, The University of New South Wales<46>
2. Taxon Based Information for GIS, Nancy Morin, Missouri Botanical Garden
3. Plant Genetic Resource Collections: an Opportunity for the Evolution of Global Data Sets, Stephanie L. Greene, USDA, Thomas Hart, Spatial Data Associates

LAND SURFACE CHARACTERIZATION: HYDROLOGY - SESSION 10

Chair, Kris Verdin, U.S. Geological Survey

1. Approaches to Automated Water Table Mapping, Kris C. Matson, North Carolina State University; and John E. Fels, North Carolina State University<121>
2. Deliniation of Chemical Hydrological Response Units (CHRUs) within a GIS Hydrochemical Modeling in the Mesoscale Broel Catchment in Germany, Ulrike Bende-Michl, University of Jena, Germany (abstract only)
3. Temporal and Spatial Aggregation of NEXRAD Rainfall Estimates on Distributed Storm Runoff Simulation, Baxter E. Vieux, University of Oklahoma, and Nadim S. Farajalla, Stone Environmental<57>
4. A Cognitively-based Approach for Hydrogeomorphic Land Classification using Digital Terrain Models, John E. Fels, North Carolina State University; and Kris C. Matson, North Carolina State University<53>

DISCRETIZATION OF SPACE - SESSION 11

Chair, Karen Kemp, NCGIA

1. Automated Grid Generation from Models of Complex Geologic Structure and Stratigraphy, Carl W. Gable, Los Alamos National Laboratory; Harold Trease, Los Alamos National Laboratory; and Terry Cherry, Los Alamos National Laboratory<66>
2. Improving the Spatial Extension of Point Data by changing the Data Model, Brian Lees, Australian National University<107>
3. Converting Administrative Data to a Continuous Field on a Sphere, Waldo Tobler, University of California, Santa Barbara<169>
4. Dynamic Finite Difference Grid Generation for Environmental Decision Support Systems, Steven P. Frysinger, James Madison University<64>

LARGE DATABASE ISSUES - SESSION 12

Chair, Jim Frew, University of California, Santa Barbara

1. Issues Linked to Geographical Information Systems in Global Environmental Research: Data Base Handling and Multi-Sensor Data Fusion, Catherine Gautier, University of California, Santa Barbara; and Pete Peterson, University of California, Santa Barbara
2. Evaluation of North and South America AVHRR 1-km Data for Global Environmental Modeling, Limin Yang, University of Nebraska-Lincoln; Zhi-Liang Zhu, Hughes STX; Jorge A. Izaurralde, National Univerity of Cordoba, Argentina; James W. Merchant, University of Nebraska-Lincoln<193>
3. Integration of GIS with Other Software Systems: Integration versus Interconnection, D.C.L. Lam, Environment Canada; D.A. Swayne, University of Guelph; C.I. Mayfield, University of Waterloo; and D.D. Cowan, University of Waterloo<164>
4. EOSDIS Data Models and Example of Implementations, Liping Di, Hughes STX; R. Suresh, Hughes STX; Doug Ilg, Hughes STX; and Ted Meyers, NASA Goddard Space Flight Center<44>

LAND SURFACE CHARACTERIZATION: VEGETATION 2 - SESSION 13

Chair, William K. Michener, Joseph W. Jones Ecological Research Center

1. Identification and Assessment of Natural Disturbances in Forested Ecosystems: The Role of GIS and Remote Sensing, William K. Michener, Jones Ecological Research Center; and Paula F. Houhoulis, Jones Ecological Research Center
2. Integrating GIS and Remote Sensing to Produce Regional Vegetation Databases: Attributes Related to Environmental Modeling, Janet Franklin, San Diego State University; and John Stephenson, Cleveland National Forest, USDA Forest Service<62>
3. Integrating Stratified Sampling, Canonical Correspondence Analysis, and GIS for Predictive Vegetation Modeling in the Spring Mts. of Southern Nevada, Andrew D. Weiss, Stuart B. Weiss, and Alisya T. Galo, Stanford University, Jan Nachlinger and Daniel Pritchett, The Nature Conservancy (abstract only)
4. Estimating Spatial Uncertainty as a Function of Scale: Implications for Landscape Ecology, Carolyn Hunsaker, Oak Ridge National Laboratory; Charles Ehlschlaeger, University of Cincinnati; Frank Davis, University of California Santa Barbara; and Michael F. Goodchild, NCGIA (abstract only)

EOSDIS Special Session

Chair, John E. Estes, University of California, Santa Barbara

1. EOS Potential User Model Development, John E. Estes, University of California, Santa Barbara<207>
2. Overview and Current Status of EOS and EOSDIS, Dixon Butler, NASA Headquarters
3. EOSDIS Commercial Applications, Michael Lawless, University of California, Santa Barbara (abstract only)

TUESDAY - January 23, 1996

Plenaries

1. GIS and Hydrologic Modeling - an Assessment of Progress, David R. Maidment, University of Texas, Austin
2. Regional Analysis of the Central Grasslands: GIS-Facilitated Pattern Analyis and Simulation, Ingrid Burke, Colorado State University

SURFACE WATER HYDROLOGY AND EROSION - SESSION 1

Chair, Fiona Ellis, Australian National University

1. Estimation and Evaluation of Spatially Distributed Model Parameters Using the Modular Modeling Systems (MMS), G.H. Leavesley, R.J. Viger, S.L. Markstrom, and M.S.

Brewer, USGS (abstract only)
2. The Application of Machine Learning Techniques to Erosion Modelling, Fiona Ellis, Australian National University<50>
3. Topography-Based Hydrological Modeling in the Elbe Drainage Basin, D.-I. Mueller-Wohlfeil, Potsdam Institute for Climate Impact Research; W. Lahmer, Potsdam Institute for Climate Impact Research; V. Krysanova, Potsdam Institute for Climate Impact Research; and A. Becker, Potsdam Institute for Climate Impact Research<102>
4. GIS Applications for Watershed Management, Yuri Gorokhovich, Earth & Environmental Sciences Program CUNY; and Lorraine L. Janus, New York City Department of Environmental Protection<204>

GIS AND HYDROLOGIC MODELING OF SNOW COVER - SESSION 2

Chair, Thomas R. Carroll, National Operational Hydrologic Remote Sensing Center, NOAA

1. Snow Estimation and Updating System (SEUS), Randy Hills, National Weather Service; Ann McManamon, National Weather Service; and Robert K. Hartman, National Weather Service
2. Spatial Distribution of Snow Water Equivalent Observations in Mountainous Terrain, Robert K. Hartman, National Weather Service; Andrew A. Rost, National Weather Service; Donald M. Anderson, National Weather Service
3. Operational Processing of Multi-Source Snow Data, Robert K. Hartman, National Weather Service; Andrew A. Rost, National Weather Service; Donald M. Anderson, National Weather Service<24>
4. Mirror-Image Round Robin Spatial Data Partitioning: A Case Study with Parallel SEUS, A.A. El Haddi, Office of Hydrology, NWS & University of Minnesota; S. Shekhar, University of Minnesota; R. Hills, Office of Hydrology, NWS; A. McManamon, Office of Hydrology, NWS

BIODIVERSITY - SESSION 3

Chair, Richard Church, University of California Santa Barbara

1. The Role of GIS and Environmental Modelling in the Conservation of Biodiversity, Brendan G. Mackey, Australian National University<116>
2. Planning Management Activities to Protect Biodiversity with a GIS and an Integrated Optimization Model, Richard Church, D. Stoms, F. Davis, B.J. Okin, University of California, Santa Barbara
3. Macroecological Studies of Species Composition, Habitat and Biodiversity Using GIS and Canonical Correspondence Analysis, L. Edward Harvey, University of Auckland<80>
4. Measuring and Modeling (Bio)Diversity: an Approach Based on Geographic, Taxonomic and Environmental Relations, Richard J. Aspinall and Diane M. Pearson, Macauley Land Use Research Institute, Aberdeen, Scotland, and Julia A. Miller, Australian National University (abstract only)

LANDSCAPE PATTERN AND CHANGE - SESSION 4

Chair, Timothy Kittel, National Center for Atmospheric Research

1. Ecosystem Modeling of Spatially Explicit Land Surface Changes for Climate and Global Change Analysis, Rebecca Mckeown, Dennis S. Ojima, Colorado State University; T.G.F. Kittel, D.S. Schimel, W.J. Parton, and T. Painter
2. Modeling Land-Cover Change From Measures of Spatial Landscape Structure, Miles G. Logsdon, University of Washington<114>
3. Mapping for Germplasm Collections: Site Selection and Attribution, Thomas Hart, Spatial Data Associates; Stephanie Greene, USDA; and Alexander Afonin, Vavilov Plant Industry Institute
4. Ecological Modelling in GIS, Peter van Horssen, University of Utrecht<203>

CONTAMINATION AND HEALTH - SESSION 5

Chair, Gerard Rushton, University of Iowa

1. Geologic Modeling for Landfill Screening: Integrating GIS with Geospatial Modeling, Christopher S. McGarry, Illinois State Geological Survey<123>
2. Groundwater Monitoring in the Alluvial Aquifer of the River Sieg, Germany - An Application of MODFLOW/MODPATH combined with GIS Analysis, Christian Michl, Friedrich-Schiller-University (abstract only)
3. Investigating the Spatial Patterns of Disease with Variable Spatial Filters, Gerard Rushton, University of Iowa

SWAT MODELING - SESSION 6

Chair, Valentina Krysanova, Potsdam Institute for Climate Impact Research

1. An Integrated Interface System to Couple the SWAT Model and Arc/Info, Ling Bian, State University of New York; Hao Sun, Leica Inc.; Clayton Blodgett, University of Kansas; Stephen Egbert, Kansas Applied Remote Sensing Program; WeiPing Li, SAI Software Consultant, Inc.; LiMei Ran, ManTech Environmental Technology Inc.; and Antonis Koussis, National Observatory of Athens<16>
2. Mesoscale Integrated Modelling of Hydrology and Water Quality with GIS Interface, Valentina Krysanova, Potsdam Institute for Climate Impact Research; Dirk-Ingmar Muller-Wohfeil, Potsdam Institute for Climate Impact Research; Alfred Becker, Potsdam Institute for Climate Impact Research<101>
3. Effect of Spatial Variability on Basin Scale Modeling, Sudhakar Mamillapalli, R. Srinivasan, J.G. Arnold, and Bernard A. Engel, Purdue University<119>
4. Modeling Wister Lake Watershed Using a GIS-Linked Basin-Scale Hydrologic/Water Quality Model, Tharacad S. Ramanarayanan, Blackland Research Center; Raghavan Srinivasan, Blackland Research Center; and Jeffrey G. Arnold, USDA-ARS<153>

ECOSYSTEM MODELING - SESSION 7

Chair, John Yarie, University of Alaska

1. Redefining the Spatial Support of Environmental Data in the Regional HydroEcological Simulation System, Jennifer L. Dungan, JCWS, Inc.; and Joseph C. Coughlan, NASA Ames Research Center<36>
2. Integration of Satellite Data and Model Simulations in a GIS for Monitoring Regional Evaporation and Biomass Production, Stephan J. Maas, USDA-ARS Shafter Research Station; and Paul C. Doraiswamy<115>
3. Using GIS to Enable Diagnostic Interaction with a Spatially Distributed Biogeochemistry Model, Rusty F. Dodson, ManTech Environmental Research Services Corp.; David P. Turner, ManTech Environmental Research Services Corp.<49>
4. A Forest Ecosystem Dynamics Model Integrated within a GIS, John Yarie, University of Alaska Fairbanks<194>

PLANT AND ANIMAL DISPERSAL - SESSION 8

Chair, Eric Gustafson, U.S. Forest Service

1. Dispersal and Mortality in a Heterogenous Landscape Matrix, Eric J. Gustafson, USDA Forest Service; and Robert H. Gardner, University of Maryland<76>
2. Integrating Ecological Tools with Remotely Sensed Data: Modeling Animal Dispersal on Complex Landscapes, Gillian Bowser, National Park Service & University of Missouri-St. Louis<211>
3. Modeling Spatial Effects of Landscape Pattern on the Spread of Airborne Fungal Disease in Simulated Agricultural Landscapes, Fred C. Bogs, Institute of Applied Sciences, University of North Texas; James M. Newell, Institute of Applied Sciences, University of North Texas; and Jeffrey W. Fitzgerald, Center for Spatial Analysis and Mapping, University of North Texas<140>
4. Landscape-Level Modeling of Spruce Seedfall Using a Geographic Information System, T. Scott Rupp, University of Alaska<157>

CLIMATIC AND ATMOSPHERIC MODELING - SESSION 9

Chair, Michael Hutchinson, Australian National University

1. The Effects of Elevation Data Representation on Mesoscale Atmospheric Model Simulations , Hoyt Walker, J.M. Leone, J. Kim, Lawrence Livermore National Laboratory
2. Biophysical Data Integration for Terrestrial Ecosystem Simulation from Watershed to Basin Scales, Robert G. Kremer, Colorado State University; R.B. Lammers, Colorado State University; M. Hartman, Colorado State University; J.S. Baron, Colorado State University (abstract only)<99>
3. An Integration of a Surface Energy Balance Climate Model with TIN and GRID in GIS, Lin Wu, California State Polytechnic University, Pomona<186>
4. Potential of GIS and Coupled GIS/Conventional Systems to Model Acid Deposition of Sulphur Dioxide, U. Dragosits, The University of Edinburgh; C.J. Place, The University of Edinburgh; and R.I. Smith, Institute of Terrestrial Ecology<48>

FISH AND WILDLIFE HABITAT - SESSION 10

Chair, Janet Franklin, San Diego State University

1. Spatial Modeling of Aquatic Habitat from a Fish's Perspective, John K. Horne, J. Michael Jech and Stephen B. Brandt, Great Lakes Center, Buffalo State College
2. Phenology Models in Complex Terrain, Stuart B. Weiss and Andrew D. Weiss, Stanford University (abstract only)
3. Spatial Modeling of Instream Biotic Integrity and Riparian Ecotone Conditions in the Big Darby Creek, Ohio, Leslie A. Zucker and Dale A. White, The Ohio State University (abstract only)
4. Linking GIS with Models of Ecological Risk Assessment for Endangered Species, H. Resit Akcakaya, Applied Biomathematics<7>

GROUND AND SURFACE WATER - SESSION 11

Chair, Miguel Acevedo, University of North Texas

1. Application of GIS Linked Environment Models over a Large Area, Thomas W. Charnock, Aston University; John Elgy, Aston University; and Peter D. Hedges, Aston University<27>
2. Rapid Appraisal of Groundwater Discharge Using Fuzzy Logic and Topography, Shawn Laffan, Australian National University<104>
3. Assessing Pollutant Loading to Bayou Chico, Florida by Integrating an Urban Stormwater Runoff and Fate Model with GIS, Toar T. Schell, University of North Texas; Miguel F. Acevedo, University of North Texas; Fred C. Bogs, University of North Texas; James Newell, University of North Texas; Kenneth L. Dickson, University of North Texas; and Foster L. Mayer, Environmental Protection Agency<158>
4. Modeling Resuspension of River Sediments using ARC/INFO, Theodore A.D. Slawecki, Limno-Tech, Inc.; Ramesh K. Raghunathan, Limno-Tech, Inc.; Victor J. Bierman, Jr., Limno-Tech, Inc.; and Paul W. Rodgers, Limno-Tech, Inc.<212>

AGRICULTURE - SESSION 12

Chair, Bernie Engel, Purdue University

1. Potential for Integrated GIS-Agriculture Models for Precision Farming Systems, T.W. Goddard, Alberta Agriculture, Food and Rural Development; L.K. Kryzanowski, Alberta Agriculture, Food and Rural Development; K. Cannon, Alberta Agriculture,

Food and Rural Development; R.C. Izaurralde, University of Alberta; T.C. Martin, University of Alberta<227>
2. Predicting Spatial Distributions of Vulnerability of Indiana State Aquifer Systems to Nitrate Leaching using a GIS, Kumar C.S. Navulur, Purdue University; Bernard A. Engel, Purdue University<139>
3. Mapping CO2 Surface Flux in an Irrigated Agricultural Area, P. Vaughan, J. Simunek, D.L. Suarez, D.L. Corwin, J.D. Rhoades
4. Model Comparison of Solute Transport Models at Regional Scale, Alejandro Mateos, Palmaven S.A., Stephen D. DeGloria and R. Jeff Wagenet, Cornell University (abstract only)

Workshop

- Progress Made Since 1991 in GIS and Hydrologic Modeling, David Maidment, University of Texas, Austin

WEDNESDAY - January 24, 1996

Plenaries

1. Modelling Complex Adaptive Systems, Chris Langton, Santa Fe Institute (abstract only)
2. Directions in GIS, Michael Goodchild, NCGIA

INTEROPERABILITY - SESSION 1

Chair, Karen Kemp, NCGIA

1. Universal GIS Operations for Environmental Modeling, Jochen H. Albrecht, University of Vechta<3>
2. Virtual Data Sets - Smart Data for Environmental Applications, Andrej Vckovski, University of Zurich; and Felix Bucher, University of Zurich
3. Easing Traditional Environmental Models into GIS, Karen Kemp, NCGIA
4. The Open GIS Ajpproach to Distributed Goodata and Geoprocessing, Kenn Gardels

SPATIAL ANALYSIS, REASONING, AND LEARNING - SESSION 2

Chair, David Cowen, University of South Carolina

1. Spatial Reasoning for Environmental Impact Assessment, B. Beattie, University of Liverpool; F. Coenen, University of Liverpool; A. Hough, Environmental Advisory Unit; T.J.M. Bench-Capon, University of Liverpool; B.M. Diaz, University of Liverpool; and M.J.R. Shave, University of Liverpool<13>
2. Development and Application of Neural Networks in GRASS GIS, Ranjan Muttiah, Blackland Research Center; Raghavan Srinivasan, Blackland Research Center; Bernard Engel, Purdue University<137>
3. Using Artificial Neural Networks for Prediction of Soil Carbon Dynamics, Daniel Kimes and Elissa Levine, NASA/Goddard Space Flight Center (abstract only)
4. Controlling a GIS by the Expert System EXCEPT, Thomas Wachter-Harms, Department of Urban Ecology; and Birgit Wendholt, IBM Germany, Development Section (abstract only)
5. An Object-Oriented Framework for Spatial Analysis Using Raster and Vector Data, Sud Menon, ESRI

TEMPORAL MODELING - SESSION 3

Chair, Michael F. Goodchild, NCGIA

1. Spatio-Temporal Interpolation in Four Dimensional Coastal Process Models, Jonathan Raper, Birkbeck College, University of London; and David Livingstone, Kingston University<151>
2. An GIS-based Many-Region Disaster Preparedness Model for the United States, Sam Cole, SUNY Buffalo
3. Animation Query Language for the Visualization of Temporal Data, Bruce Rex, Pacific Northwest National Laboratory; and John Risch, Pacific Northwest National Laboratory<155>
4. Temporal GIS and Spatio-Temporal Modeling, May Yuan, University of Oklahoma

VISUALIZATION: COGNITION AND LANDSCAPE VIEWING - SESSION 4

Chair, Carol Johnston, University of Minnesota

1. Modelling Environmental Cognition of the View with GIS, Jonathan Baldwin, University of Leicester; Peter Fisher, University of Leicester; Joseph Wood, University of Leicester; and Mitchel Langford, University of Leicester<59>
2. Landscape Visualization Using DEM Data Derived from Digital Photogrammetry, David R. Miller, Macauley Land Use Research Institute<133>
3. Towards a Virtual Reality Interface for Landscape Visualization, Peter Berger, Brimsom Laboratories; Paul Meysembourg, University of Minnesota; Jim Sales, University of Minnesota; and Carol Johnston, University of Minnesota<215>
4. Quality, Visualization and Use of Terrain Models in Physical System Modelling,

Michael McCullagh, University of Nottingham, UK

LAND USE MODELING - SESSION 5

Chair, Bradley O. Parks, University of Colorado

1. The Future of Spatial Modeling for Understanding and Predicting Landscape Transformations, Robert Costanza, University of Maryland (abstract only)
2. Integrated Approach to Land Use Analysis, Dennis S. Ojima, Colorado State University; and W. J. Parton<144>
3. Spatial Resolution of Crop Models in the Estimation of Regional Agroecological Effects of Climate Change: How Fine is Fine Enough?, William E. Easterling, University of Nebraska (abstract only)
4. Integrating Spatial Models into Local Land Use Decision Making, Richard O. Flamm, Florida Marine Research Institute

MODELING ENVIRONMENTS AND LANGUAGES - SESSION 6

Chair, Thomas Maxwell, University of Maryland

1. Facilitating High Performance, Collaborative Spatial Modeling, Thomas Maxwell, University of Maryland; and Robert Costanza, University of Maryland<122>
2. Future 4-Dimensional Hydrological Modeling Environments, Paul Houser, University of Arizona
3. A Spatial Modelling Language that Unifies Dynamic Environmental Models and GIS, Cees G. Wesseling, Utrecht University; Willem P.A. van Deursen, Resource Analysis; Peter A. Burrough, Utrecht University<179>
4. A Framework for Integrating Environmental Models to Simulate Forest Ecosystem Dynamics, Robert G. Knox, NASA/Goddard Space Flight Center; Virginia Kalb, NASA/Goddard Space Flight Center; Elissa R. Levine, NASA/Goddard Space Flight Center; and Uday Bindingnavle, Hughes STX Corporation

CELLULAR MODELING - SESSION 7

Chair, James D. Westervelt, University of Illinois, Urbana-Champaign

1. Facilitating Mobile Objects within the Context of Simulated Landscape Processes, James D. Westervelt, Construction Engineering Research Laboratories; and Lewis D. Hopkins, University of Illinois, Urbana-Champaign<188>

2. SELES: A spatially Explicit Landscape Event Simulator, Joseph Fall, Andrew Fall, Simon Fraser University, Canada
3. Methods and Techniques for Rigorous Calibration of a Cellular Automaton Model of Urban Growth, Keith C. Clarke, Stacy Hoppen, and Leonard Gaydos
4. Some Guidelines for Implementing Spatially Explicit, Individual-Based Ecological Models within Location-Based Raster GIS, Roger L. Slothower, Cornell University; Paul A. Schwarz, Cornell Theory Center; and Kevin M. Johnston, Environmental Systems Research Institute, Inc.<159>

VISUALIZATION: INTEGRATION - SESSION 8

Chair, Mark Gahegan, Curtin University

1. Visualization Strategies for Exploratory Spatial Analysis, Mark Gahegan, Curtin University<67>
2. Combining GIS Based Environmental Modeling and Visualization: Another Window on the Modeling Process, Ian D. Bishop, The University of Melbourne; Chris Karadaglis, The University of Melbourne<17>
3. Efficient Data Exchange: Integrating a Vector GIS with an Object-Oriented, 3-D Visualization System, James Kuiper, Argonne National Laboratory; A. Ayers, Argonne National Laboratory; R. Johnson, Argonne National Laboratory; M. Tolbert-Smith, U.S. Department of Energy<103>
4. Visualization of Historical Wildfire Data: Application of a DX-Oracle Interface, Trish Duce, University of Montana; Ellen Voth, University of Montana; and Patricia Andrews, Intermountain Research Station, USDA Forest Service<202>

INDIVIDUAL BASED MODELING - SESSION 9

Chair, Chris Langton, Santa Fe Institute

1. Individual-Based Models in Ecology: An Overview, Donald L. DeAngelis, National Biological Service & University of Miami; and D.M. Fleming, National Biological Service; L.J. Gross, University of Tennessee; and W.F. Wolff, Forschungszentrum Juelich (abstract only)
2. Agent-Based Modeling of Prehistoric Settlement Systems in the Northern American Southwest, Timothy A. Kohler, Washington State University; Carla R. Van West; Eric P. Carr; and Christopher G. Langton
3. Emergence and the Simulation of Transportation Systems, Christopher L. Barrett, Los Alamos National Laboratory
4. Modeling Ecological Patterns and Processes Using Agent-Based Simulations and GIS, James H. Brown, University of New Mexico (abstract only)

MULTIPLE CRITERIA/SOLUTIONS AND OPTIMIZATION - SESSION 10

Chair, Gunnar Misund, SINTEF Informatics

1. Solving the Clear-Cut Scheduling Problem with Geographic Information Technology and Constraint Reasoning, Gunnar Misund, SINTEF Informatics; Bjørn Sigurd Johansen, SINTEF Informatics; and Geir Hasle, SINTEF Informatics<135>
2. Interacting Fields Approach for Evolving Spatial Phenomena: Application to Erosion Simulation for Optimized Land Use, Lubos Mitas, National Center for Supercomputing Applications, University of Illinois; Helena Mitasova, University of Illinois; William M. Brown, University of Illinois; Mark Astley, University of Illinois<136>
3. Modelling and Supporting Multi-Actor Spatial Planning Using Multi-Agents Systems, Nils Ferrand, Leibniz Lab, France<60>
4. Optimal Field Management for Regional Water Quality Planning, Babak Negahban, Breedlove, Dennis & Associates; C.B. Moss, J.W. Jones, J. Zhang, W.D. Boggess and K.L. Campbell (abstract only)

COLLABORATION - SESSION 11

Chair, Steve Carver, University of Leeds

1. Environmental Modeling and Collaborative Spatial Decision-Making: Some Thoughts and Experiences Arising from the I-17 Meeting, Steve Carver, University of Leeds; Steve Frysinger, James Madison University; and Rene Reitsma, CADSWES, University of Colorado
2. Agent Mediated Consensus-Building for Environmental Problems: A Genetic Algorithm Approach, David A. Bennett, Southern Illinois University; Marc P. Armstrong, University of Iowa; and Greg A. Wade, Southern Illinois University<15>
3. Collaborative GIS in Ecosystem Management System, Jeff Wang, Scientific Visualization Center/Lockheed Martin Service Group, US EPA/National Environmental Supercomputing Center (abstract only)
4. Client-Server Approaches to Model Integration within GIS, Raghbir Sandhu, University College London; and Philip Treleaven, University College London<160>

3-DIMENSIONS AND DATA MODELS - SESSION 12

Chair, Claudia Faunt, U.S. Geological Survey

1. Applications of 3D Delaunay Triangulation Algorithms in Geoscientific Modeling,

Roberto Lattuada, Institute for Animal Health; Jonathan Raper, Birkbeck College

2. Hybrid Data Structures Based on Octree and Delaunay Tetrahedral Tessellations (DTT) in 3-D GIS, Quingquan Li and D. Li, Wuhan Technical University, China (abstract only)
3. Construction and Role of 3D Geological Framework Models, Claudia C. Faunt, USGS and A. Keith Turner, Colorado School of Mines (abstract only)
4. Topologic and Hierarchical Spatial Object Models for Database Generalization, Martien Molenaar

THURSDAY - January 25, 1996

Plenary

- Progress and Future Directions - Panel Session
  - Peter Burrough, University of Utrecht
  - Michael Goodchild, University of California, Santa Barbara
  - David Maidment, University of Texas, Austin
  - Michael Hutchinson, Australian National University
  - Robert Coulson, Texas A&M University

  Moderator, Carol Johnston, University of Minnesota

Panel

- Teaching GIS for Environmental Modeling
  - Peter Burrough, University of Utrecht, Netherlands
  - David Maidment, University of Texas, Austin
  - Jonathan Raper, Birkbeck College, UK
  - Stephen Kessel, Curtin University, Austin

  Organizer, Karen Kemp, NCGIA, University of California Santa Barbara

Go to Main Menu

Go to Poster Presentations

NCGIA National Center for Geographic Information and Analysis

Third International Conference/Workshop on Integrating GIS and Environmental Modeling

January 21-25, 1996
Santa Fe, New Mexico, USA

Poster Presentations

A Prototype of a Landscape Forest Ecosystem Management Tool Using a State Transition Model and a GIS

- presented by Magdiel Ablan
Institute of Applied Sciences, University of North Texas
Denton, TX, USA
with Susan Monteleone, Miguel F. Acevedo
email: ablan@unt.edu

Forest Fire Modeling in the Swiss National Park

- presented by Britta Allgöwer
Department of Geography, University of Zurich
Zürich, Switzerland
with Reto Schöning (University of Zurich)
email: britta@gis.geogr.unizh.ch

A Methodology to Build a Classification of Spatialized and

Multivariate Data

**- presented by Michel Arnaud
CIRAD
Montpellier, Cedex 5, France
with Jean Pichot
email: arnaud@cirad.fr**

Problems of Multi-resolution Integration in Dynamic Simulation

**- presented by George Ball
University of Arizona, School of Renewable Natural Resources
Tuscon, Arizona 85721
with Bernard P. Zeigler (Electrical and Computer Engineering, University of Arizona), Richard Schlichting (Computer Sciences, University of Arizona), Michael Marefat Electrical and Computer Engineering, University of Arizona), and D. Phillip Guertin (School of Renewable Natural Resources, University of Arizona)
email: gball@nexus.srnr.arizona.edu**

Production of Regional Maps of Long-Term Runoff Using Simple GIS-Based Methods

**- presented by Gary Bishop
Ogden Professional Services
Corvallis, OR, USA
with M. Robbins Church
email: bishop@mail.cor.epa.gov**

System Integration of GIS and Environmental Models in the Personal Computer Environment

- presented by W. G. Booty
National Water Research Institute, Environment Canada, Burlington, Ontario, Canada, L7R 4A6
with D.C.L. Lam (National Water Research Institute), D.A. Swayne (Computing and Information Science, University of Guelph), C.I. Mayfield (Environmental Biology, University of Waterloo), L. Leon Vizcaino (Civil Engineering, University of Waterloo), G.S. Bowen (Ontario Ministry of Environment and Energy, Toronto), I. Wong and D.F. Kay (National Water Research Institute, Environment Canada)
email: bill.booty@cciw.ca

Spatial Models Highlight Radiological Hazards

- presented by Patrick J. Bresnahan
University of South Carolina / U.S. DOE SRTC
Columbia, SC, USA
with Cowen/Jensen (USC), Mackey (SRTC)
email: pat@otis.cla.sc.edu

Spatially Explicit Desert Tortoise Population Model

- presented by Douglas R. Briggs
Department of Computer Science, University of Illinois at Urbana-Champaign
Champaign, IL, USA
with James Westervelt (CERL), Shawn Levi (Dept. of Geog, UIUC), Steve Harper (Dept. of Ecology, Ethology & Evolution, UIUC)
email: douglas@gis.uiuc.edu

Generation and Utilization of DEM Data for Environmental

Restoration

- presented by Greg Cole
Earth & Environmental Science Division, Los Alamos National Laboratory
Los Alamos, NM, USA
email: gcole@lanl.gov

Integrated Risk & Impact Assessment Modeling a Support of Sustainable Development in an Estuarine Environment

- presented by David Cowen
Department of Geography, Uninversity of South Carolina
Columbia, SC, USA
with Dwayne Porter (Belle Baruch Marine Lab, University of SC)
email: cowend@garnet.cla.sc.edu

Spatio-Temporal Object Handling to Model the Effects of Acid Deposition

- presented by Ferko Csillag
Geography, University of Toronto
Mississauga, ON, Canada
with Scott Mitchell, Rebecca Handcock (Dept. of Geography, University of Toronto)
and Charles Driscoll, Kiran Sequeira (Dept. of Civil and Environmental Engineering, Syracuse University)
and Hon Yau (Northeast Parallel Architecture Center, Syracuse University)
email: fcs@geog.utoronto.ca

Characterization of the recharge and discharge components of the Death Valley regional ground-water flow system using

remote sensing and GIS techniques

- presented by Frank A. D'Agnese
U.S. Geological Survey
Denver, CO 80225
with Claudia C. Faunt (U.S. Geological Survey, Denver, Colorado) and A. Keith Turner (Colorado School of Mines, Golden, Colorado)

Quality Assurance in GIS and Environmental Modelling Applied to an Example of Phosphate Saturated Soils

- presented by Marien de Bakker
Environmental Sciences, Van Hall Institute
Groningen, The Netherlands
email: marien.deBakker@vhall.nl

Digital Elevation Data and GIS Projects

- presented by Robert de Sawal
National Mapping Division, U.S. Geological Survey
Denver, CO, USA
email: rfdesawal@usgs.gov

Application of CAD Framework Techniques to Systems Integration in Environmental Modelling

- presented by Jonathan Deckmyn
CRS 4, Center for Advanced Studies Research and Development in Sardinia
Cagliari, Italy
with Sally Kleinfeldt (CRS4), Claudio Paniconi (CRS4), Pieter van der Wolf (Delft University of Technology-DIMES), and Olav ten Bosch (Delft University of Technology-DIMES)

email: jdeckmyn@crs4.it

A Biogeographic Exploration of the Relationship Between Vegetation Distribution and Enviromental Variables in Wyoming, USA

- presented by Kenneth L. Driese
Department of Botany, University of Wyoming
Laramie, WY, USA
with William A. Reiners (Botany, UWYO)
email: gap@botsun2.uwyo.edu

Dynamic Linkages of GIS and a Coupled Hydro-Geomorphic Model CLAWS

- presented by Jinfan Duan
Oregon State University
Corvallis, OR, USA
with Gordon E. Grant and Chaur-Fong Chen
email: duan@fsl.orst.edu

The Tyranny of Scale and the Multifractal Paradigm

- presented by Ralph Dubayah
Geography, University of Maryland
College Park, MD, USA
email: rdubayah@geog.umd.edu

Protection Afforded Land-Cover and Terrestrial Vertebrate Diversity in Utah

- presented by Thomas C. Edwards, Jr.
Utah Cooperative Fish & Wildlife Research Unit, National Biological Service
Logan, UT, USA

with Scott D. Bassett
email: tce@nr.usu.edu

A GIS Decision Support System for Fire and Alien Weed Management for the Nature Reserves of South Africa: Spatial Simulation to Application

- presented by Dean Fairbanks
Natural Resources and Development Program
Division of Water, Environment and Forestry (Environmentek) CSIR
Pretoria, South Africa
with Brian W. Van Wilger, David H. McKelly, Katherine Reast, David H. Le Maitre
email: dfairban@csir.co.za

Natural Resources Planning by Way of a Geographic Spreadsheet Modeling Approach

- presented by Dean Fairbanks
Natural Resources and Development Program
Division of Water, Environment and Forestry (Environmentek) CSIR
Pretoria, South Africa
with Mike P. Adam
email: dfairban@csir.co.za

The Response of Vegetation to Change of Annual Rainfall in the Sahel Region of Africa, and Its Dependence on Soil Type

- presented by George W. Fisher
Earth and Planetary Sciences, Johns Hopkins University
Baltimore, MD, USA
with Elissa Levine (NASA GSFC Code 923)

email: gfisher@jhu.edu

GIS Approaches to Targeted Siting of Riparian Buffer Strips: Trade-offs Between Realism and Complexity

- presented by Jeremy Fried
Department of Forestry
Michigan State University
126 Natural Resources
East Lansing, MI, USA
with Mark Zweifler, Michael Gold, Daniel Brown
email: jeremy@msu.edu

Using Remote Sensing Analysis of Landsat Data to Evaluate an Integrated Socio-Economic Model of Deforestation in the Amazon

- presented by Robert Frohn
Geography, Remote Sensing Research Unit, UCSB
Santa Barbara, CA, USA
with Ken McGwire (Desert Research Institute), John E. Estes (RSRU), Virginia H. Dale (Oak Ridge Natl. Lab)
email: frohn@geog.ucsb.edu

Watershed Management Division Support System

- presented by Chris Fulcher
Agriculture Economy, CARES
Columbia, MO, USA
with Yan Zhan, Tony Prato
email: fulcher@cares.missouri.edu

Metadata and Standards - Communicating Between Disciplines in the Encounter of GIS and Environmental

Modeling

- presented by David T. Hansen
MPGIS, U.S. Bureau of Reclamation U.S. Department of the Interior
Sacramento, CA, USA
email: dhansen@mpgis7.mp.usbr.gov

The Building of a GIS Interface to Modflow Utilizing ARC/INFO GRID 7.0 and Its Application to a USGS Groundwater Model in California's San Joaquin Valley

- presented by Tom Heinzer
U.S.B.R. MPGIS Service Center
Sacramento, CA, USA
with Mike Sebhat (GIS Analyst), William Greer (Hydrologist)
email: msebhat@mpgis1.mp.usbr.gov

Incorporating Expert Opinion in Modelling Wildlife Species Distributions

- presented by Allan D. Hollander
Dept. of Geography, Biogeography Lab, UCSB
Santa Barbara, CA, USA
email: adh@geog.ucsb.edu

Modeling Present and Potential Future Tree Importance Values in the Eastern United States

- presented by Louis R. Iverson
Northeastern Forest Experiment Station
359 Main Road
Delaware, Ohio 43015
Voice: 614-368-0097

Fax: 614-368-0152
with Anantha M. Prasad
email: iverson@trees.neusfs4153.gov

A GIS Database and Its Use to Quantify Nitrogen Retention by Natural Wetlands

- presented by Asa Jansson
Department of Systems Ecology
The Beijer International Institute of Ecological Economics / Stockholm University
Stockholm, Sweden
with Carl Folke, Sindre Langaas (UNEP/Grid Arendal, c/o Dept. of Systems Ecology)
email: asaa@beijer.kv.se

What are the Educational Requirements for Integrating GIS and Environmental Modeling?

- presented by Steve Kessell
Department of GIS, School of Computing, Curtin University of Technology
Perth, Western Australia
email: kessell@cs.curtin.edu.au

Environmental Modeling and Prognosis System as a Functional Subsystem of Ecological Information Management System of the Republic of Bashkortostan

- presented by Rustem Z. Khamitov
Ministry of Emergency Situations and Environmental Safety
12/1, 3 Marta St.
450005, Ufa, Bashkortostan

Tel(fax): (3472)28-75-90 with Vladimir E. Gvosdev, Sergei V. Pavlov, and Andrei N. Yasiliev (Institute for Problems of Applied Ecology and Natural Resources Use)
email: root@ippeprb.bashkiria.su

Using GIS and Coupled Models for Understanding Forest Ecosystem Dynamics

- presented by Elissa R. Levine
Biospheric Sciences Branch, NASA / Goddard Space Flight Center
Greenbelt, MD, USA
with R. G. Knox, K. J. Ranson, J. A. Smith, N. Chauhan (GWU, NASA Goddard), D. L. Williams, J. F. Weishanpel (Univ. of Central Florida), G. Sun (SSAI, NASA), A. D. Friend (IVE-Edinburgh Research Station, Peniwik, Scotland), S. Fifer (Hughes STX, NASA)
email: elissa@lichen.gsfc.nasa.gov

GISMO: On Linking the EPIC Simulation Model with GIS

- presented by Tim Martin
Department of Renewable Resources, University of Alberta
Edmonton, AB, Canada
with H. Neiman
email: tim.martin@ualberta.ca

Creation of a Managed Areas GIS Database of the Conterminous United States for use in Ecosystem Analysis of Managed Versus Unmanaged Areas

- presented by R. Gavin McGhie
Geography, Remote Sensing Research Unit (RSRU),

UCSB
Santa Barbara, CA, USA
with Karen Kline, John E. Estes
email: gavin@geog.ucsb.edu

Analyzing Thematic Map Accuracy Using Generalized Linear Mixed Models

- presented by Gretchen G. Moisen
USDA Forest Service Intermountain Research Station
Ogden, UT, USA
with D. Richard Cutler (Utah State, Dept. of Mathematics and Statistics), Thomas C. Edwards, Jr. (USDI National Biological Service)
email: moisen@edumath.math.usu.edu

Dynamic-Stochastic Model of Ground Water Balance for River Basins

- presented by Igor S. Pashkovsky
Department of Hydrogeology
Moscow State University
Moscow, 119899 RUSSIA
Phone: 095/115-9986
FAX: 095/115-9992
with Egenea Yu. Potapova and Boris A. Shmagin
email: shmag@hydro.geol.msu.su (Shmagin's)

The Clear Box Image Processing Simulator

- presented by Micha Pazner
Geography, University of Western Ontario
London, ON, Canada
with Brian Reynolds (student, U. of Western Ontario)

email: pazner@ssol.uwo.ca

Natural Resources Research Institute & Brimson Laboratories

- presented by Jim Sales
CWE, NRRI
Duluth, MN, USA
email: jsales@sparkie.nrri.umn.edu

Exploratory Visualization of Environmental Data

- presented by Eva-Maria Stephan
Department of Geography, University of Zürich
Zürich, Switzerland
email: stephan@gis.geogr.unizh.ch

Selecting Biodiversity Management Areas in the Sierra Nevada Region

- presented by David M. Stoms
Biogeography Lab, Institute for Computational Earth System Science, UCSB
Santa Barbara, CA, USA
with Frank Davis, Richard Church (Geog, UCSB), B. J. Okin (Geog, UCSB), Joe Walsh
email: stoms@geog.ucsb.edu

Integrating Novel Applications Into a GIS Framework: Some Examples

- presented by Dave Swayne
Computer and Information Science, University of Guelph
Guelph, ON, Canada
with D.C.L Lam, J.D. MacNeil, Adrian Harding, Mark

Mayo, Alex Storey, Ken Brown, and Doug Kay
email: dswayne@snowhite.cis.uoguelph.ca

Merging Two Large Mineral Location Databases Using Logistic Regression

- presented by Stella W. Todd
Management Assistance Corporation of America
Ft. Collins, CO, USA
with Deborah J. Shields, Douglas D. Brown
email: stellat@lamar.colostate.edu

GIS-Based Modeling of Desert Tortoise Habitat in the Mojave Desert

- presented by Joseph M. Watts
Remote Sensing Technology Division, U.S. Army Topographic Engineering Center
Alexandria, VA, USA
email: watts@tec.army.mil

Visual Representation and Analysis of the Climatic Data Using GIS

- presented by Harumi Kitajima Yanagimachi
Faculty of Economics, Shinshu University
3-1-1 Asahi, Matsumoto, Nagano 390 Japan
Phone: +81-263-35-4600 ex.3335
Fax: +81-263-36-7220
with Kazutaka Iwasaki (Shizuoka University) and Kenji Sato (Pasco Corp.)
email: yanagi@econ.shinshu-u.ac.jp

Creation of a 3D Perspective Classified Forest Map Using

Geographic Information and Remote Sensing Integration

- presented by H. Yildirim

TUBITAK-MAM Space Technologies Department

P.O. Box 21 41470 Gebze-Kocaeli/TURKEY

with E. Alparslan, B. Bilge, H. Kurar, O. Divan, and S. Elitas

Go to Main Menu

Go to Program Proceedings

Originally created by Karen Kline, enhanced and revised by Chris Stebbins.

Third International Conference/Workshop on Integrating GIS and Environmental Modeling

Santa Fe, New Mexico
January 21-25, 1996

LIST OF SESSIONS AND PROCEEDINGS

- Sunday Workshops, January 21, 1996
- Monday, January 22, 1996
    - Plenaries
    - Climate/Weather Interpolation
    - Internet/Data Access 1
    - Land Surface Characterization: Vegetation 1
    - International Agricultural Development
    - Remote Sensing and Image Processing
    - Internet/Data Access 2
    - Land Surface Characterization: Soils
    - Land Surface Characterization: DEMs
    - Data Integration/New Data Sources
    - Land Surface Characterization: Hydrology
    - Discretization of Space
    - Large Database Issues
    - Land Surface Characterization: Vegetation 2
    - EOSDIS Special Session
- Tuesday, January 23, 1996
    - Plenaries
    - Surface Water Hydrology and Erosion
    - GIS and Hydrologic Modelling of Snow Cover
    - Biodiversity
    - Landscape Pattern and Change
    - Contamination and Health
    - SWAT Modeling
    - Ecosystem Modeling
    - Plant and Animal Dispersal
    - Climatic and Atmospheric Modeling
    - Fish and Wildlife Habitat
    - Ground and Surface Water
    - Agriculture

- Wednesday, January 24, 1996
    - Plenaries
    - Interoperability
    - Spatial Analysis, Reasoning, and Learning
    - Temporal Modeling
    - Visualization: Cognition and Landscape Viewing
    - Land Use Modeling
    - Modeling Environments and Languages
    - Cellular Modeling
    - Visualization: Integration
    - Individual Based Modeling
    - Multiple Criteria/Solutions and Optimization
    - Collaboration
    - 3-Dimensions and Data Models
- Thursday, January 25, 1996

Go to the Main Menu

NCGIA National Center for Geographic Information and Analysis

SANTA FE CONFERENCE CONTRIBUTION INDEX

This list includes both head and contributing authors.

Letter Search: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A

B

- Baker R.H.A.
- Baldwin John
- Ball George
- Ball J.E.
- Baron J.S.
- Barringer James
- Bassett Scott
- Beard Kate 1
- Beard Kate 2
- Beattie Bridget
- Becker Alfred 1
- Becker Alfred 2
- Bench-Capon T.J.M
- Bende-Michl Ulrike (abstract)
- Benedict Karl (abstract)
- Bennett David
- Berger Peter
- Bian Ling
- Bierman V.J.
- Bilge B. (abstract)
- Bindingnavle U.
- Bishop Gary
- Bishop Ian
- Boggess W.D. (abstract)
- Bogs Fred C. 1
- Bogs Fred C. 2
- Bogs Fred C. 3
  (abstract)
- Bolstad Paul V.
- Booth Trevor
- Booty W.G. (abstract)
- Bowen G.S. (abstract)
- Bowser Gillian
- Brandt Stephen B.
- Bree David S.
- Bresnahan Patrick
- Brewer M.S. (abstract)
- Briggs Douglas (abstract)
- Brown Dan G. (abstract)
- Brown Douglas D.
- Brown H. James (abstract)
- Brown Ken
- Brown William M.
- Bucher Felix
- Burrough Peter A.
- Buttenfield Barbara

C

D

E

F

G

H

I

J

K

L

M

N

O

P

- Prasad Anantha M.
- Prato Tony (abstract)
- Pritchett D. (abstract)

Q

R

- Ramanarayanan Tharacad
- Ramlal Bheshem
- Ran LiMei
- Ranson K.J. (abstract)
- Raper Jonathan
- Raper Jonathan
- Reitsma Rene
- Rex Bruce
- Reynolds Keith (abstract)
- Reynolds Brian
- Rhoades J.D.
- Risch John
- Rodgers Paul W.
- Rogowski Andrew
- Rost Andrew A. 1
- Rost Andrew A. 2
- Rupp Scott

S

- Sales Jim
- Sandhu Raghbir
- Sato Kenji
- Schell Toar 1
- Schell Toar 2 (abstract)
- Schimel D.S.
- Schlichting Richard
- Schoning Reto
- Schwarz Paul A.
- Sear David (abstract)
- Sebhat Michael
- Sequeira Kiran
- Sersland C.A.
- Shave M.J.R.
- Shekhar Sashi
- Shields Deborah

T

U

V

W

Go Back to Top of Page

Go to Main Menu

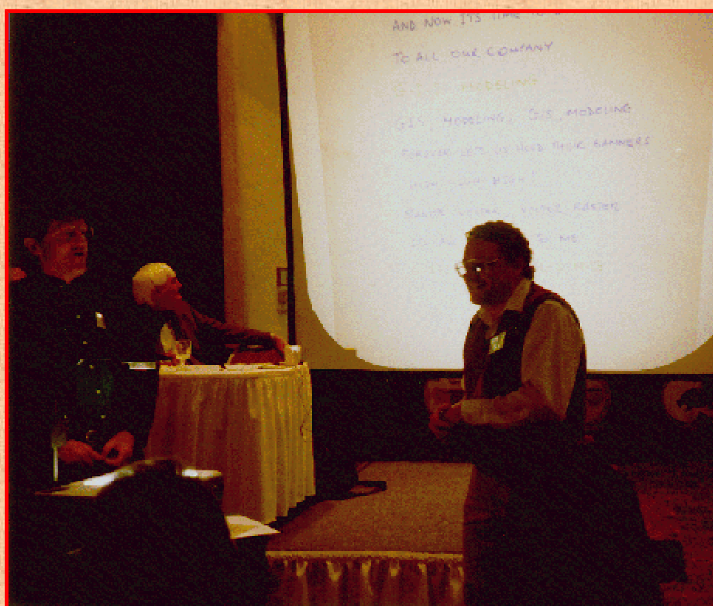
Go to Program Proceedings

Go to Poster Presentations

Go to the Main Menu

NCGIA National Center for Geographic Information and Analysis

The Third International Conference/Workshop on Integrating GIS and Environmental Modeling



Go to the Main Menu...

The Third International Conference/Workshop on Integrating GIS and Environmental Modeling was held under the auspices of the U.S. National Center for Geographic Information and Analysis in Santa Fe, NM from Sunday January 21 to Thursday January 25, 1996. The conference follows previous meetings in Boulder, Colorado in 1991 and Breckenridge, Colorado in 1993. Total attendance at the three conferences exceeded 1700.

The papers from the previous two meetings were published as edited books:

```
Environmental Modeling with GIS, edited by M.F. Goodchild,
B.O. Parks, and L.T. Steyaert.  New York: Oxford University
Press (1993)

GIS and Environmental Modeling, edited by M.F. Goodchild,
L.T. Steyaert, and B.O. Parks, et al..  Fort Collins:
GIS World (1996).
```

For the 1996 conference, much use was made of the World Wide Web

(http://www.ncgia.ucsb.edu) to distribute information prior to the conference, and to give broader access to the conference papers. Rather than an edited volume, therefore, the organizers decided that it would be appropriate to publish the proceedings in the form of a CD; this would also have the advantage of cutting the time between conference and publication significantly.

Like the Web site, the CD has been organized as a complete record of the conference, including program, the full text of papers, abstracts of posters and workshops, and the list of attendees. The papers are published as submitted, and have not been edited to conform to any standard style.

Standards for citation of electronic material are still in a state of flux; meanwhile, we recommend the use of either of the following forms:

> Author (1996) Title. In Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, NM, January 21-26, 1996. Santa Barbara, CA: National Center for Geographic Information and Analysis. CD.

or:

> Author (1996) Title. In Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, NM, January 21-26, 1996. Santa Barbara, CA: National Center for Geographic Information and Analysis. http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/main.html.

Many people helped to organize the conference, and to assemble the proceedings. Special thanks go to NCGIA staff members: Sandi Glendinning, Karen Kemp, Karen Kline, Chris Stebbins, Terry Figel, LaNell Lucius, Elan Sutton and Ken Cushing; and to the members of the conference organizing committee: Louis Steyaert, USGS; Bradley Parks, University of Colorado; Michael Crane, USGS; Carol Johnston, University of Minnesota; John Wilson, Montana State University; Denice Shaw, US EPA; and Sandi Glendinning, NCGIA. Also gratefully acknowledged are the generous contributions from several organizations including National Science Foundation (NSF), US Geological Survey, Environmental Systems Research Institute (ESRI), National Oceanic and Atmospheric Administration (NOAA), GIS World and Los Alamos National Laboratory.

Copies of the CD are available from the NCGIA Publications Office, 3510 Phelps Hall, University of California, Santa Barbara, CA 93106-4060, USA, email: ncgiapub@ncgia.ucsb.edu, phone: 805 893 8224.

Go to the Main Menu...

NATIONAL SCIENCE FOUNDATION

The NCGIA is supported by a major grant from the National Science Foundation (SBR-88-10917). This support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or of the NCGIA.

Student Scholarships

Third International Conference/Workshop on Integrating GIS and Environmental Modeling

January 21-25, 1996, Santa Fe, New Mexico, USA

- Chris Brown, San Diego State University
- Catherine Dibble, University of California, Santa Barbara
- Ulrike Dragosits, University of Edinburgh
- Nils Ferrand, France
- Mark Forney, University of California, Berkeley
- Matt Gardner, East Anglia University
- Prasanna Gowda, Ohio State University
- Violet Gray, University of California, Santa Barbara
- Jim Hipple, University of Utah
- Allan Hollander, University of California, Santa Barbara
- Shawn Laffan, Australian National University
- Bryan Mark, Syracuse University
- Pawel Mizgalewicz, University of Texas, Austin
- BJ Okin, University of California, Santa Barbara
- Michael Parke, University of Hawaii
- Jeffrey Piampiano, University of Maine
- Skip Repetto, Montana State University
- Bill Starmer, University of California, Santa Barbara
- Sara Stillman, Montana State University
- James Tsai, Georgia Tech
- Hailing Wang, North Carolina State University

Go to Main Menu.

TERRAIN ANALYSIS WORKSHOP

Organized by John P. Wilson, Department of Earth Sciences, Montana State University and John C. Gallant, Centre for Resource and Environmental Studies, Australian National University.

Part 1. Terrain Analysis Programs

DEM Data Sources and Interpolation Methods by Michael F. Hutchinson, Centre for Resource and Environmental Studies, Australian National University.

Generation of Primary Topographic Attributes with TAPES-C and TAPES-G by John C. Gallant, Centre for Resource and Environmental Studies, Australian National University.

Generation of Secondary Topographic Attributes with DYNWET, EROS, SRAD and WET by John P. Wilson, Department of Earth Sciences, Montana State University.

Linkages between TAPES-C and ARC/INFO by Jinfan Duan, Department of Forest Engineering, Oregon State University and Gordon E. Grant, Pacific Northwest Research Station, Corvallis, Oregon.

Part 2. Soil-Landform Applications

Landform Classification for Soil-Landscape Studies by Barbara J. Irwin and Stephen J. Ventura, Institute for Environmental Studies and Department of Soil Science, University of Wisconsin-Madison.

A Soil-Terrain Model for Estimating Spatial Patterns of Soil Organic Carbon by Jay C. Bell, D.F. Grigal, Department of Land, Water, and Climate, University of Minnesota and P.C. Bates, Department of Natural Resources, Western Carolina University.

Terrain Analysis for New Methods of Soil Mapping by Neil J. McKenzie, Paul E. Gessler, CSIRO Division of Soils, and Philip J. Ryan, CSIRO Division of Forestry, Canberra, Australia.

Identification of Areas Experiencing Net Erosion and Deposition by Skip Repetto and John P. Wilson, Department of Earth Sciences, Montana State University.

Part 3. Hydrologic Applications

Mapping Contributing Areas for Stormwater Discharge to Streams by Mark O. Zweifler, Department of Forestry, Daniel G. Brown, Department of Geography, Jeremy S. Fried, and Michael A. Gold, Department of Forestry, Michigan State University.

Topography and GIS-Based Hydrological Modeling in the Elbe Drainage Basin by Dirk I.

Muller-Wohlfeil, W. Lahmer, and Valentina Krysanova, Potsdam Institute for Climate Impact Research, Potsdam, Germany.

Soil Moisture Modeling in Humid Mountainous Landscapes by J. Alan Yeakley, Department of Environmental Sciences and Resources, Portland State University, Paul V. Bolstad, Department of Forest Resources, University of Minnesota, Wayne T. Swank and James M. Vose, Coweeta Hydrologic Laboratory, USDA-Forest Service, Otto, North Carolina.

Stochastic Analysis of a Coupled Surface/Subsurface Hydrologic Model by Gregory M. Pohll and J.J. Warwick, Graduate Program of Hydrologic Sciences, University of Nevada-Reno.

Part 4. Ecological Applications

Terrain Variables Used for Predictive Mapping of Vegetation Communities in Southern California by Janet Franklin, Department of Geography, San Diego State University, Paul McCullough, San Diego Data Processing Corporation, and Curtis Gray, Department of Geography, San Diego State University.

Mapping Existing Vegetation in the Northern Great Plains Using Satellite Imagery and Terrain Attributes by Jonathan M. Wheatley, John P. Wilson, Department of Earth Sciences, Montana State University, Zhenqui Ma and Roland L. Redmond, Wildlife Spatial Analysis Laboratory, University of Montana.

Applications of Terrain Analysis Methods to Boreal Forest Ecosystems by Brendan G. Mackey, Department of Geography, Australian National University, Daniel W. McKenney, K. Baldwin, R. Sims, Ontario Department of Natural Resources, and John C. Gallant, Centre for Resource and Environmental Studies, Australian National University.

Workshop on Network-Accessible Data Repositories

NCGIA Conference on GIS and Environmental Modelling

Santa Fe, New Mexico - January 21- 25, 1996

Workshop Organizers
Ray Ford, Mike Sweet, Ron Righter, Joe Glassy
University of Montana
Missoula, Montana 59812, USA

Call for Workshop Participation

As part of the NCGIA 3rd International Conference/Workshop on GIS and Environmental Modelling in Santa Fe, New Mexico January 21-25, a group of us have been asked to organize a 2 hour workshop that addresses the major technical issues involved in constructing a network accessible information repository containing GIS and ecosystem information. The workshop is scheduled for Sunday afternoon, January 21st (the day before the conference general sessions begin). Our goal is to attract both presenters and discussion participants. The target is to have 6 - 8 presenters, each providing a short (10 minute) presentation on specific technical aspects of repository construction from the perspective of someone actively building such a repository. These brief presentations would form the basis for follow-up discussion and Q/A participation by the larger workshop audience. We want the workshop presenters to focus on technical issues involved in repository construction, rather than on a "show and tell" about their repository contents. We are particularly interested in projects that are not specifically World-Wide-Web based, but that have a Web component or connection. That is, it is relatively easy these days to find information that addresses the construction of a Web site; what is more difficult is determining how to add Web links or other forms of "Web-accessibility" to an existing or customized repository.

Though the focus of the workshop is to be on technical issues, rather than on the contents of specific repositories, one of the benefits that we expect each presenter to gain from workshop participation is visibility for his/her repository. In that regard, we've been tentatively promised a time slot and room where participants (presenters or other interested parties) can set up demo's of their repositories to allow workshop participants and other conference attendees to engage in more detailed follow up on the technical details AND contents of their repositories. We're hoping to attract outside support that allows us to provide a network connection, but we can't count on being able to provide a large cluster of workstations and a robust high-speed network connection. Instead, what we suggest is that participants who are interested and for whom this is feasible arrange to bring a "portable" running a local copy (or subset) of their repository. Our experience in doing repository demo's suggests that this is more effective because it is less prone to network delays and the problems inherent in trying to setup demo's on borrowed equipment.

What we are trying to do now is to collect a list of people who are interested in being considered as a workshop presenter, a non-presenting workshop participant, and a repository demonstrator. We've identified you as someone active or knowledgable in the technical aspects of repository construction AND also a possible attendee at the Santa Fe conference. What we are looking for now is an indication of your potential interest in the workshop activies, or a referral to other people or groups you think might be interested. What we ask is the following.

- (a) **Potential Presenters** If you or someone from your lab/organization is interested in being considered as a workshop presenter, please contact Ray Ford, via e-mail "ford@cs.umt.edu", phone at 406-243-2964, or FAX at 406-243-4076. We're particularly interested in the technical focus of your project. For potential demo's, we're interested in the contents of your repository and how you'd plan to present your demo (what you would bring and what you would need).
- (b) **Information/Suggestions** If you have questions about participation, suggestions on topics the workshop might address, or simply want additional information, contact Mike Sweet, via e-mail at "sweet@selway.umt.edu", phone at 406-243-5265, or FAX at 406-243-5265. Any information you could provide about the referral would be appreciated.

Tentative List of Participant Presentations

The Development and Application of a GIS-Based Prehistoric Resource Distribution Model in Archaeological Research

This workshop illustrates an application of GIS-based paleoenvironmental models in archaeological research. The workshop emphasizes the process of model development and comparison of the resulting resource distribution with a known archaeological site distribution. The presentation proceeds through all the major steps of model development: problem definition, data needs assessment and data acquisition, entry and preprocessing of data, development of a data management system, manipulation and analysis of the resulting resource distribution, and generation of output maps and other products. Throughout the workshop, budgetary and time requirements for each step are provided, allowing for an illustration of these requirements in the context of model development and utilization. Overall, the process outlined in this workshop yields a reconstruction of past environmental conditions in a spatial framework, allowing for a more refined analysis of human land use than has historically been available in archaeological research.

Participants:
Karl Benedict, Univerity of New Mexico, Albuquerque, NM
Louis Scuderi, University of New Mexico, Albuquerque, NM
Richard Watson, San Juan College, Farmington, NM

Bryan C. Pijanowski and Bradley Parks, Organizers

Land Transformation/Land Use Modeling Workshop

Preliminary Abstract

This land use/land transformation modeling workshop will focus on three main themes: (1) the use of landscape ecology approaches for land transformation/land use change modeling, and, in particular, the need to address the integration of the human dimensi ons and aspects of scale into modeling; (2) computational strategies to spatial land use modeling; and (3) public reaction and involvement in issues related to land transformation modeling. This workshop is integrated into the land use modeling paper ses sion (Session 5 - Wednesday) held on Wednesday afternoon from 1:30 to 3:00pm. Workshop participants are welcome to provide additional insights/experiences related to the three broad themes of the workshop and paper session.

Bryan Pijanowski, of Michigan State University, will begin the workshop by providing an overview of the breadth of approaches that each workshop and paper presenter will cover. He will then provide an overview of the land transformation modeling project he helped to direct for the past two years. This model, which is called the Land Transformation Model (LTM), is a relative risk, spatial, numerical model developed using ARC/INFO GRID. Its purpose is to forecast land use change in Michigan's Saginaw Ba y Watershed over the next fifty years and to provide an assessment of the effects of these changes on the environment and the quality of life for the residents of the watershed.

In a project very similar to Pijanowski's, Keith Clarke of Hunter College in New York City, will discuss his involvement in the Human Induced Land Transformation Modeling project being sponsored by the USGS Global Change Project. Keith will also provide an overview of his experiences on interacting with the media and the public on this high visibility project.

Thomas Maxwell of the University of Maryland Institute for Ecological Economics will discuss distributed, spatial modeling computational needs and software architectural requirements.

George Ball, of the University of Arizona, will round out the workshop with a presentation on tradeoffs between scale and resolution of modeling, computational requirements and the accuracy of predictions.

After the workshop, a discussion, involving the workshop audience, workshop and paper presenters for the related paper session will be lead by Pijanowski. A comparison of the approaches and experiences of all workshop participants is anticipated.

Bryan C. Pijanowski
Department of Entomology

Michigan State University
East Lansing, Michigan 48824
office: (517) 432-0039
lab : (517) 355-4561
fax : (517) 353-4354
email: pijanows@esalsun10.ent.msu.edu

Brad Parks
Cooperative Institute for Research in Environmental Sciences (CIRES)
Campus Box 449
University of Colorado
Boulder, Colordao 80309-0449
phone: 303-497-6330
fax: 303-497-6513
email: 74063.3124@compuserve.com

Stacy Hoppen, Keith C. Clarke, Leonard J. Gaydos, and William Acevedo

Communicating Scientific Findings to the General Public

Abstract

The Human Induced Land Transformation project (HILT), part of the U.S. Geological Survey Global Change Research Program, has generated considerable public interest during the first two phases of the project: the data assembly and visualization phase carried out for the San Francisco Bay area; and the model building and calibration phase. The third phase, now under way for the Washington/Baltimore region, is a large scale test of the model's portability to another urban area. The modeling community has been kept informed of progress and findings of the HILT project through conventional academic outlets such as conference and journal publications as well as through more popular outlets such as newspaper, television, and a World Wide Web site. The program, sample data and documentation of the model will be available for downloading to those with an Internet connection. While recognizing that not all research has the same degree of public appeal as this project, additional effort was invested to establish contacts, publicize, and distribute the historical visualizations and model predictions generated by this work in the media, publishing, and in education. In this paper, we will cover the lessons learned in publicizing HILT, the techniques used and the criteria evolved for successful involvement of the general public in the modeling of land cover transitions.

Introduction

During the 1994 Association of American Geographers (AAG) meeting in San Francisco, a local TV news anchorman stepped from behind his desk and stood in front of a map. He proceeded to show Bay Area viewers how their piece of the world was transformed by human activity since the Gold Rush. An animation of land use change filled the screen, much like a weather map. The anchorman showed how the settlements grew from several red dots in 1850. As the years clicked away on the map he showed how the Bay Area grew. "We approach the 60's where you'll see a sudden surge here in the South Bay with the Silicon Valley exploding, literally exploding!" (Wilson, 1994).

A serious researcher's aversion to media attention is certainly understandable. The press has a tendency for superficial coverage, suggesting exposure that will trivialize or misrepresent one's work. Scientists frequently imagine their work to be of most interest to a narrow audience, relying on conference presentations and journal articles to publicize their results, and the advantages of media exposure are often overlooked. Publicity emphasizes the relevance of one's work, increasing chances for future funding, it reaches people outside the scientific community who may be interested in the work, and it can inform and influence public opinion about local policy issues.

This television spot, one of several, was the outcome of a successful data visualization of a local issue that the public could identify with: the dramatic population growth in the San Francisco Bay area. The Human Induced Land Transformations project (HILT), under the U.S. Geological Survey Global Change research program, developed an urban growth model for predicting land cover changes into the 21st century, specifically the conversion of rural land to urban land. The project involved three components: model development, calibration, and predictions. While the model's predictions were always expected to be of interest to the general public, what caught media attention in 1994 was a historical animation of urban expansion assembled by William Acevedo (Acevedo and Bell, 1994; Bell, et al., 1995). The animation was created using the temporal database for the San Francisco/Sacramento area that was compiled for calibrating the model. The animation took on a life of its own, a surprising outcome since it was originally thought of as an intermediate step in the process of model calibration. Even without conclusive results there was a great response from the media, providing a lesson in how timely topics and strong graphics can both excite and inform the public through the mass media.

Planning

Thorough planning is an essential aspect of publicity. The media blitz is intense and brief and may be a unique chance for promotion. An organized strategy optimizes this opportunity to solidly position one's work, to reach the widest audience possible, and to lay the groundwork for further opportunities after the initial event.

Find a Hook

Essential to successful media attention is finding a "hook" that will convince a journalist that your work has enough appeal to justify news coverage. A hook should focus on an object or an event rather than an abstract idea. Local topics revolving around a current policy debate have a strong appeal to local newscasters. Modern media is a highly visual medium so a hook can be as simple as an arresting graphic to illustrate the results. As well as a hook, a packaged story needs to make sense of the facts by putting them into context, such as the historical significance or an explanation of cause and effect.

Part of the hook for the HILT project was that it was a local research team studying a local phenomenon. The research team includes a group from the U.S. Geological Survey who are based at NASA Ames Research Center at the southern end of San Francisco Bay. Since the initial data set used to test the model was for the San Francisco Bay area, the animation of urban growth was pertinent for local residents. The topic has broader environmental implications as well: the contribution of urbanization to global warming. Both of these elements were incorporated into the press release to present the model as newsworthy on several levels.

Attracting Media Attention

Television and newspapers have daily story slots to fill, and because of staff and time constraints, much of the news centers around fields that provide good press materials: politics and entertainment. If your story is not the usual media fare but comes pre-packaged, it will add a welcome diversity to the news. Of course it is easier to get news time if one is located in a city with good media coverage.

Publicity about the HILT model was scheduled during a week in which two conferences were being held in San Francisco, The Association of American Geographers and the Fifth Global Warming Conference. Papers about the development of the model were presented at both. The press release was tied to a session at the AAG conference that featured four papers about the model. The AAG press room at the convention hotel supplied copies of our press release, and we followed up on the inquires for more information that were generated. Our press release included a brief background about the project, quotable sentences, and contact information stating a time, date, and a place at which representatives would be available to discuss the project. Keep the language to easily understandable terms. Journalists have a good general background but may be unprepared for technical jargon. Given too much information, the journalist will further refine and repackage the story, possibly adding a different slant to the material. By presenting the journalist with an already pared-down explanation of the research allows one more control over the final story. A USGS press agent was available for advice. If you don't have a publicist in your organization, consider the expertise available to you via colleagues or company staff persons.

Anticipating possible questions and formulating their answers is an important part of preparing to meet the press. Reporters prefer to let their source explain and interpret the significance of the facts, partly in an effort to remain objective and partly limited by time constraints. Prepared answers will guarantee that you communicate your main points concisely. Appearing as an authoritative source, whether by education, title, or institution will help convince the public, your final audience, of the validity of your findings. While it is essential to be available at the times stated by the press release, one must also be willing to be flexible about schedules since reporters juggle many stories as they approach their deadline. If possible, it is better to schedule press events in the morning because it allows time for the story and footage to be assembled.

Materials for Release

A press release, color prints of graphics, digital graphics, and a videotape of an animation of historical growth were released to the press. The historical animation received the most attention, serving as a visualization tool for understanding past growth in the area.

Besides understanding what materials make a good story, it is important to know the appropriate format for these materials. Text, graphics, and films may be required to be in a form different from what you are accustomed to or are able to produce with your software. All the materials submitted should be clearly labeled so that they are readily understandable even if they are shown out of context.

Reaction to HILT Publicity

Although some local interest was expected in this animation, we were not prepared for the level of response it received. After a press release from the USGS, a call came to the AAG press room from Channel 7 requesting a copy of the animation. This was followed by calls from the San Francisco Chronicle and the San Jose Mercury News.

Reporters from the Chronicle and Mercury News greeted us the morning of our presentations at the AAG meeting. They watched with interest when the animation played on a monitor. After asking many questions, it became obvious that they thought the animation tracing Bay Area growth was newsworthy. The Channel 7 crew came in to conduct its interview and to take away a betacam tape copy. The newspaper reporters asked for digital files of the base year maps used to construct the animation. While we presented our papers the Channel 5 crew showed up for an interview.

The Bay Area woke up the next morning to six color maps of the Bay Area on the front page of the San Francisco Chronicle with the headline "The Spreading Sprawl" (Petit, 1994). The night before, both Channels 7 and 5 aired the interviews taped during the AAG meeting. The remaining local TV stations called Friday requesting tapes as well. They were picked up by Channels 4 and 2 and appeared on the news that night.

The San Jose Mercury News carried a banner headline, "Mapping the Mega Sprawl," in the weekly Science and Medicine Section (Chui, 1994) the following week on the same day that the Greenbelt Alliance held a press conference releasing a new report on land conservation and development trends using the USGS animation as background. Those who watched TV that night saw the moving map again, in reference to Greenbelt Alliance's message. This coverage came as international delegates at the Fifth Global Warming Conference in San Francisco viewed the animation as part of a presented paper (Kirtland, et al., 1994).

Television coverage produced a variety of stories built around this single animation, ranging from running the animation as a story in itself to using it as a supporting fact in the Greenbelt Alliance report on urbanization. Most newscasters used the animation to make the point that urbanization is occurring more rapidly than realized and to discuss the implications of this for the future.

Follow Up

Even after your fifteen minutes of fame, the opportunities for publicity continue. The bulk of the two pages of contacts generated by this single press conference trickled in for several weeks after the initial event. Some of these contacts led to information exchange between county planning departments and the HILT research team, contact with others researching the same topic, and a television interview. William Acevedo was invited, as a local newsmaker, to tell the inside story of how the animated map was made (Chu, 1994).

Internet Publicity

Increasingly important outlets for information are home pages available through web browsing software. They are accessible sources of information for those with Internet access, centralized, easily linked to other locations, and allow instantaneous update. A HILT web page has been maintained from the beginning as a way for the research team to keep up with the separate phases of the project. The web site has also become the spot for the public to find out about the project, access user documentation, and to find links to other relevant home pages. Visits to web sites increase dramatically as links are built between home pages. A hot link between our home page and a paper by Michael Batty, known for his work on the simulation of urban forms, significantly increased the traffic to our site (Batty, 1995).

Conclusion

It was exciting to see our work displayed with such enthusiasm by the media. We feel there are lessons we can all learn from this experience. The public is interested in what we have to say when we say it right, and the press can be convinced that research produces worthwhile stories. TV stations and newspapers sought us out after we issued a simple press release. People were fascinated by seeing their region grow through time, and public awareness of the pace of change and its effects were heightened.

We all have some talent for interpreting complex data and making it understandable. Current data processing and visualization software provide superb graphics tools. Aside from the boost it can give your research, publicity makes research relevant by helping the public understand the context and processes of sometimes overwhelming changes taking place in the world. Our story didn't just list names and places, but illustrated dynamic changes and implied relationships over time.

It is hoped that our experience will encourage other scientists who would like to publicize their findings. Although not all discoveries will attract broad public interest, there is probably more curiosity about their work than most researchers suspect. Even without previous experience with the media, a relevant topic and strong graphical presentation can lead to a successful promotion campaign.

References

Journals:

Clarke, K.C., Gaydos, L., Hoppen, S., (1996) "A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area," Environment and Planning B (in press).

Kirtland D., DeCola L., Gaydos L., Acevedo W., Clarke K., Bell C., (1994) "An Analysis of Human-Induced Land Transformations in the San Francisco Bay/Sacramento Area," World Resource Review: vol. 6(2); pp 206-217.

Conference Presentations and Publications:

Acevedo, W., Bell, C., (1994) "Time Series Animation of Historical Urban Growth for the San Francisco Bay Region," Abstracts of the Association of American Geographers 90th Annual Meeting. San Francisco.

Bell, C., Acevedo, W., Buchanan, J., (1995) "Dynamic Mapping of Urban Regions: Growth of the San Francisco/Sacramento Region," Proceedings, Urban and Regional Information Systems Association. San Antonio, pp 723-734.

Clarke, K.C., Hoppen, S., Gaydos, L.J., (1996) "Methods and Techniques for Rigorous Calibration of a Cellular Automaton Model of Urban Growth," Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21-25, 1996. Santa Barbara: National Center for Geographic Information and Analysis. WWW and CD.

Clarke, K.C., (1994) "A Cellular Automaton Model of Urban Growth in the San Francisco Bay Area," Abstracts of the Association of American Geographers 90th Annual Meeting. San Francisco.

De Cola, L. (1994) "Exploratory Analysis of Human-Induced Land Transformations in Central California," Abstracts of the Association of American Geographers 90th Annual Meeting. San Francisco.

Gaydos L. J., Acevedo, W., Bell, C. (1995) "Using Animated Cartography to Illustrate Global Change," Proceedings, International Cartographic Association, Barcelona; pp 1174-1178.

Gaydos, L., (1994) "Golden Gate to Golden Foothills: California Urbanization as an Example of Human-Induced Land Transformation," Abstracts of the Association of American Geographers 90th Annual Meeting. San Francisco.

Hoppen, S., Clarke, K.C., Gaydos, L., (1995) "Calibration of a Cellular Automaton Urban Growth Model," Abstracts of the Association of American Geographers 91st Annual Meeting. Chicago.

Other Publications:

(1994) "Geographers Track Bay Area Urbanization," GIS World; June; p 16.

Chui, G., (1994) "Mapping the Mega Sprawl," San Jose Mercury News, April 5; pp 1E+.

Gaydos, L. J. (1996) "Beyond the Weather Map," Geo Info Systems (in press).

Petit, C., (1994) "The Spreading Sprawl; Thumbs Up for Video on Bay Area Growth," San Francisco Chronicle, April 1; pp A1+.

Television:

Chu, R., (1994) Bay Area People, Interview KTVU, Channel 2, Oakland, May 8.

Garcia, A., (1994) Evening news broadcast, ABC, Channel 7, San Francisco, March 31.

Hanamura, W., (1994) Evening news broadcast, CBS, Channel 5, San Francisco, April 5.

Jones, M., (1994) Evening news broadcast, NBC, Channel 4, San Francisco, April 5.

Schaub, J., (1994) Evening news broadcast, CBS, Channel 5, San Francisco, March 31.

Corral, E., (1994) Evening news broadcast, KTVU, Channel 2, Oakland, April 1.

Wilson, P., (1994) Evening news broadcast, NBC, Channel 4, San Francisco, April 1.

Internet:

Human Induced Land Transformations: http://geo.arc.nasa.gov/usgs/HILTStart

Temporal Urban Mapping: http://edcwww.cr.usgs.gov/umap/umap.html

Batty, M., (1995) "The Computable City," Keynote Address: Fourth International Conference on Computers in Urban Planning and Urban Management, Melbourne, Australia, July 11-14. http://www.geog.buffalo.edu/Geo666/batty/melbourne.html

---

Stacy Hoppen, Graduate Student, Department of Geology and Geography, Hunter College-City University of New York.

Keith C. Clarke, Department of Geology and Geography, Hunter College-City University of New York, and the City University of New York Graduate School and University Center.

Leonard J. Gaydos, US Geological Survey, EROS Data Center, NASA�Ames Research Center, Moffett Field, CA.

William Acevedo, US Geological Survey, EROS Data Center, NASA�Ames Research Center, Moffett Field, CA.

The Swarm Modelling System

Chris Langton, Santa Fe Institute

Swarm is being developed as an object-oriented software "laboratory" for the study of multi-agent systems. The objective of Swarm is to provide researchers with much of the software apparatus typically employed in a computer simulation of multiple-agents interacting in the context of a dynamic environment. Technical details of Swarm will be explained and there will be opportunities for experimenting with Swarm applications.

NEW RESEARCH FRONTIERS IN 3D SUBSURFACE CHARACTERIZATION

This workshop was proposed so that develops and users of 3D modeling software can demonstrate and discuss their approaches and applications to modeling the environment in 3 dimensions. The workshop is organized to allow direct interaction between presenters and attendees during scheduled open discussion periods.

Workshop Coordinators -

Frank A D'Agnese, Claudia C. Faunt (U.S. Geological Survey, Denver, Colorado) and A. Keith Turner (Colorado School of Mines, Golden, Colorado)

12:00 Greeting - Workshop Coordinators

12:05 Recent Research Trends - Keith Turner

12:20 Presentation: "A geocellular approach to environmental modeling in the subsurface"

Thomas R. Fisher, Senior Product Geoscientist,
Zycor-Landmark Graphics, Austin, Texas

1:00 Open Discussions

1:15 Presentation: "The practical realities of 3D modeling in the geosciences"

Simon Houlding, Chairman and Director of Technical Services,
Lynx Geosystems, Vancouver, British Columbia, Canada

2:00 Open Discussions

2:15 Presentation: "Voxel models of geology in engineering"

Bogdan Orlic
ITC - International Institute for Aerospace Survey and Earth Sciences,
Delft, The Netherlands

3:00 Open Discussions

3:15 Presentation and Demonstration: "Riding the WINDOWS technology wave with volume analysis"

Eric Bayer, Senior System Consultant,
Intergraph Corporation, Lafayette, Colorado

4:00 Open Discussions

4:15 Closing

GIS, Knowledge Bases, Groupware, and Decision Support: Linked Technologies for Resource Management

Using a hands-on approach, this workshop will utilize groupware technology in an interactive workshop covering the role of technologies such as Geographic Information Systems, Groupware, Knowledge-based systems and decision support technology in support of environmental modeling and resource management. Workshop leaders will guide the group through real-world applications of these technologies and facilitate on-line exercises designed to broaden the mind and illustrate the potential for linked GIS-technologies. Examples from federal and local experiences will be used to illustrate the potential for collaborative applications of GIS integrated with these other technologies.

Workshop Leaders:

Brenda Faber, PhD, Consortium for International Earth Science Information Network

John Calkins, Environmental Systems Research Institute, Inc.

Keith Reynolds, PhD, USDA Forest Service, Pacific Northwest Forest and Range Experiment Station

On the use of GIS in mapping spatio-temporal patterns of environmetnal variability

Assaf Anyamba

Workshop abstract:

A fundamental problem in global environmental change studies is in the understanding of spatio-temporal patterns of climate variability from time series data. Climate variability occurs at a range of time scales and has different effects over different areas. In order to be able to understand these patterns of variation and their causal mechanisms requires techniques that can decompose time series data into a limited and manageable number of patterns of variability. This workshop will explore the use of a GIS spatio-temporal procedure based on standardized principal components analysis in the extraction of patterns of climate variability from Normalized Difference Vegetation Index (NDVI) time series data set for Africa. We shall illustrate how patterns of variation in NDVI on seasonal to interannual time scale relate to meso-scale patterns of variation in the atmosphere - ocean system. Information will be provided on how this technique can be applied to other areas of the world that are sensitive to interannual climate variability and in exploration of other problems related to global environmental change.

The Clark Labs for Cartographic Technology and Geographic Analysis
Clark University, Worcester, MA 01610

Multi-Criteria Modelling in GIS using Fuzzy Measures

J. Ronald Eastman and Hong Jiang

Abstract:

This workshop explores procedures for multi-criteria modelling in GIS in the context of fuzzy measures. Fuzzy measures (such as the possibilities of Fuzzy Set Theory and the beliefs of Dempster-Shafer Theory) are continuous expressions of set membership that are particularly useful in the representation of subjective knowledge and their combination in multi-criteria models. This workshop presents a live demonstration of the procedures that are used to develop the fuzzy measures, and their subsequent aggregation through the use of a special procedure known as an Ordered Weighted Average (OWA). Through parameter adjustment, the OWA operator can not only accommodate the weighting of fuzzy factors, but also allows one to specify, on a continuous basis, the specific amounts of ANDORness and tradeoff required. As a consequence, the procedure is able to evaluate not only the traditional intersection (AND) and union (OR) operations of fuzzy sets, but also a continuum of possibilities in between (including the traditional weighted linear combination of Multi-Criteria Evaluation). As will be demonstrated, the results can be radically different, depending upon the procedure chosen. The rational for various combinations of results can be radically different, depending upon the procedure chosen. The rational for various combinations of ANDness, ORness, and tradeoff is explored with the opportunity for direct participation in the specification of the models developed. Participants will all receive an advance copy of the OWA software module that can be used with the IDRISI geographic analysis system, or any similar raster package that includes an 8-bit binary image format.

Clark University
Worcester, MA 01610

Design issues for GIS-based environmental modelling: socio-economic and biophysical perspectives

Organisers : Richard Aspinall (MLURI, Aberdeen) and Brian Lees (ANU, Canberra)

This workshop develops a structured approach to resolving issues associated with organising and structuring data for GIS-based environmental models. It uses practical examples from the speakers research experiences in Australia, US and the UK.

Data quality, accuracy and sampling of spatial data (Richard Aspinall, Fiona Ellis), and time-related issues (Brian Lees, Bill Fitzgerald) are discussed and both environmental models (Larry Band, Kim Lowell) and socio-economic models (Greg Elmes) are considered. Additionally, a computer science perspective methods for supporting various types of integration is described (Mark Gahegan).

The workshop will provide a detailed overview of the many issues associated with developing environmental modelling within GIS. Those attending should gain a more integrated view of the way in which these issues can be resolved.

Proposed Speakers

Richard Aspinall : MLURI, UK
Brian Lees : ANU, Australia
Fiona Ellis : ANU, Australia
Bill Fitzgerald : ANU, Australia
Mark Gahegan : Curtin, Australia
Kim Lowell : Laval, Canada
Larry Band : Toronto, Canada
Greg Elmes : W Virginia, USA

John P. Wilson, Geographic Information and Analysis Center, Department of Earth Sciences, Montana State University, Bozeman, MT 59717-0348.

GIS-based Land Surface/Subsurface Modeling: New Potential for New Models?

ABSTRACT

Many soil erosion and non-point source pollution models have been combined with geographic information systems (GISs) to capitalize on the spatial analysis and display capabilities of these new software tools and provide regional soil erosion and non-point water quality assessments during the past decade. These models and the GIS software were developed by different groups of scientists (at different times and places) and the potential benefits and limitations of this integration warrant closer scrutiny. This paper addresses these data integration issues at two levels: (1) the input data requirements and role of GIS in providing these data for six popular land surface/subsurface models (USLE, ANSWERS, AGNPS, CMLS, LEACHM and TOPMODEL) are reviewed, and (2) the types of research that will be required to build stronger links between GIS and land surface/subsurface models and the potential for building new and improved GIS-based models in the future are examined. The effect of data resolution (i.e., the number and size of units used to represent model inputs) and input data estimation methods on model results is emphasized because: (1) the GIS-based applications vary widely in terms of the data structures and methods used to organize the spatially distributed model inputs, and (2) the development of new GIS-based methods for estimating model inputs will promote the development of new and improved environmental models.

INTRODUCTION

Mathematical models integrate existing knowledge into a logical framework of rules and relationships (Moore and Gallant 1991) and can be used to: (1) improve our understanding of environmental systems, that is, as a tool for hypothesis testing, and (2) provide a predictive tool for management (Beven 1989, Grayson et al. 1992). Many environmental models require spatially distributed inputs because solutions to accelerated soil erosion, non-point source pollution and other pervasive environmental problems involve changes in land use and management at the hillslope and catchment scales (Moore et al. 1993b). The paucity of input data at the preferred spatial resolution and difficulty of handling multiple inputs that vary in different ways across landscapes (i.e. the modifiable unit area problem) have emerged as major impediments to the successful application of models in environmental management.

Modern geographic information systems offer new opportunities for the collection, storage, analysis, and display of spatially distributed biophysical and socioeconomic data (Goodchild et al., 1993, 1996). Several soil erosion and non-point source pollution models have been modified and combined with GIS software to take advantage of these new capabilities and provide regional soil erosion and non-point water quality assessments during the past decade (e.g. Hession and Shanholtz 1988, Ventura et al. 1988, De Roo et al. 1989, Petach et al. 1991). The GIS is used to compile and organize the input data and/or display the model outputs in these applications, and the integration is achieved by passing data between the GIS and model of choice (e.g. Joao and Walsh 1992, Wilson et al. 1993) or by embedding the model in the GIS or a decision support system organized around the GIS (e.g. James and Hewitt 1992, Engel et al. 1993, Romanowicz et al. 1993). However, the GIS software, digital databases, and environmental models were developed by different groups of scientists (at different times and places) and the potential benefits and limitations of this integration warrant closer scrutiny (Moore et al. 1993b).

This paper critically reviews these data integration issues and explores the potential for building new and improved spatial models of land use systems and key environmental processes in the future. The paper is divided into two parts. The first part examines the input data requirements and role of GIS in providing these data for six popular land surface/subsurface models: the USLE (Wischmeier and Smith 1978), ANSWERS (Beasley and Huggins 1982), AGNPS (Young et al. 1987), CMLS (Nofziger and Hornsby 1986), LEACHM (Wagenet and Hutson 1989), and TOPMODEL (Beven and Kirkby 1979). The effect of data resolution (i.e. the number and size of units used to represent model inputs) and input data estimation methods on model results is emphasized because the GIS-based applications vary widely in terms of the data structures and methods used to organize the spatially distributed model inputs. The second part examines the types of research that will be required to build stronger links between GIS and land surface/subsurface models and the potential for building new and improved GIS-based models in the future.

Return to the abstract

RECENT GIS-BASED LAND SURFACE/SUBSURFACE MODELING APPLICATIONS

The descriptions of the six individual models which follow emphasize their input data requirements, how these data have been obtained or generated with a GIS, and the response of each model to variations in input data since the GIS-based applications vary widely in terms of the methods, scale, and data structures used to generate model inputs. The choice of scale is important because grid-scale parameters may not exist for some processes and they may not be related to point measurements in others (Binley et al. 1989, Goodrich and Woolhiser 1991).

The six models

USLE

The Universal Soil Loss Equation (USLE) is a simple multiplicative model that was derived from over 10,000 plot years of data (Wischmeier and Smith 1978). The factor values were recently updated following the analysis of thousands of new measurements (Renard et al. 1993) and a revised version of the model has been substituted in place of the original model for farm conservation planning in the United States (Glanz 1994). The USLE and Revised Universal Soil Loss Equation (RUSLE) can be written as:

$$A = R\,K\,L\,S\,C\,P \qquad (1)$$

where *A* is soil loss in tons per acre, *R* is the rainfall-erosivity factor, *K* is a soil erodibility factor, *L* is a slope length factor, *S* is a slope steepness factor, *C* is a cover-management factor, and *P* is a supporting practices factor. Land use and management are represented by *CP* and can be estimated from field observations or farm records (Wilson 1989, Busacca et al. 1993). *CP* may also, with some difficulty, be inferred from aerial photography or satellite imagery and ground-truth data (Stephens et al. 1985, Ventura et al. 1988, Fraser et al. 1995). Climate erosivity (*R*) can be computed directly from rainfall intensities and amounts. *R* varies on a regional scale. Soil erodibility (*K*) values have been measured or estimated for all mapped soil series as part of the county soil survey program in the United States. Soil series are mapped at scales of 1:15,000 to 1:20,000 in these surveys. The effects of topography and hydrology on soil loss are characterized by the combined *LS* factor. Soil loss predictions are more sensitive to slope steepness than slope length. Estimation of the *LS* factor poses more problems than any of the other factors in the USLE and is a particular problem in applying it to landscapes as part of a GIS (Wilson 1986, Renard et al. 1991, Moore and Wilson 1992, 1994).

Several attempts have been made to combine this model with a GIS and generate regional soil loss assessments. Hession and Shanholtz (1988) transformed the USLE into a raster-based model and combined it with the Map Analysis Package (Tomlin 1980) and a sediment delivery ratio to estimate sediment loadings to streams from agricultural land in Virginia's Chesapeake Bay. A single *R* was obtained from published maps and used for each county, *K* was obtained from county soil survey reports, *LS* was calculated for each cell by inserting slope length and the weighted cell slope into the appropriate USLE equations, *C* was determined from Landsat imagery and *P* was assumed to be constant and equal to unity. One hectare (100 m by 100 m) grid cells were used for all data except elevation. The majority rule was used to assign USLE factor values to cells for discontinuous data such as soil erodibility and the centroid value was assigned to each cell for continuous data such as the topographic factor. Elevation was sampled at a 4 ha cell resolution (200 m grid spacing) and slopes were determined by weighting the slope between each cell and its eight neighbors. The topographic factor was calculated at this coarse resolution and then interpolated to a 1 ha grid size because of (computer hardware?) cost constraints. A sediment delivery ratio was calculated for each agricultural land cell and combined with the USLE soil loss to estimate the sediment that reaches the stream.

Two other studies chose the polygon data structure of a vector GIS and treated the USLE as a zone-based model. Ventura et al. (1988) used a series of GIS polygon overlays and FORTRAN programs to estimate soil erosion in Dane County, Wisconsin. A seamless digital soil data layer for the entire county was prepared from 181 detailed soil maps and used to assign *R*, *K*, and *LS* factor values. Five land cover types were classified from a Landsat Thematic Mapper (TM) scene and combined with boundary information for Public Land Survey System (PLSS) quarter sections, incorporated areas, and wetlands to assign *C* and *P* factor values. These land cover and soil data layers were then overlaid and used to estimate soil erosion for the 500,000 polygons (0.4 ha average size) in Dane County labeled as row crop and hay/meadow cover types.

James and Hewitt (1992) used a series of ARC/INFO coverages and Arc Macro Language programs to build a decision support system for the Blackfoot River drainage in Montana. Their system was based on the Water Resources Evaluation of Nonpoint Silvicultural Sources (WRENSS) model which, in turn, incorporates a modified version of the USLE to estimate potential soil erosion. *R* was estimated from published maps and historic snow survey data, *K* values were estimated from a series of digital and paper USDA-Natural Resource Conservation Service (NRCS) and USDA-Forest Service (FS) soil survey maps, *LS* values were estimated from 3 arc-second digital elevation models (DEMs) using ARC/INFO's GRID module, and a land cover data layer was prepared from a Landsat TM scene. Some additional data processing was required because: (1) some of the soil survey source maps delineated NRCS soil series and others delineated FS land-type units at scales ranging from 1:250,000 to 1:24,000; (2) the topographic factor estimates were resampled to a larger cell size, stratified into classes, and converted into a vector format to ensure compatibility with the other model data layers; and (3) a generalized land cover data layer was generated without the benefit of extensive ground-truth data. The user interface that was developed as part of this decision support system allows data browsing and querying at the basin level and data modeling at the subwatershed level.

The GIS was used to transform the USLE into a semi-distributed model in these applications. However, there are a number of important assumptions embedded in the USLE that help to explain why the application of this model to landscapes is much more difficult than its application to soil loss plots. These key assumptions include: (1) sediment deposition (and soil losses and gains between neighboring areas) is not represented; (2) landscapes must be divided into uniform slope facets; and (3) runoff is generated uniformly over the catchment.

The first assumption represents a major practical problem because the USLE does not distinguish those parts of hillslope profiles experiencing net erosion and deposition. Cesium-137, a radioisotope by-product of atmospheric nuclear weapons testing programs, has been successfully used to document patterns of erosion and deposition in fields (Ritchie and McHenry 1990). Busacca et al. (1993), for example, used Cs-137 to measure net soil loss and gain at 143 locations in a 46 ha closed watershed in the Palouse region of northern Idaho and found that net erosion averaged 11.6 t ha^{-1} yr^{-1} from erosional areas (60% of the watershed) and that deposition averaged 18.6 t ha^{-1} yr^{-1} onto depositional areas (Figure 1). The USLE (and RUSLE) should only be applied to those parts of the landscape experiencing net erosion (Wischmeier 1976, Wilson 1986).

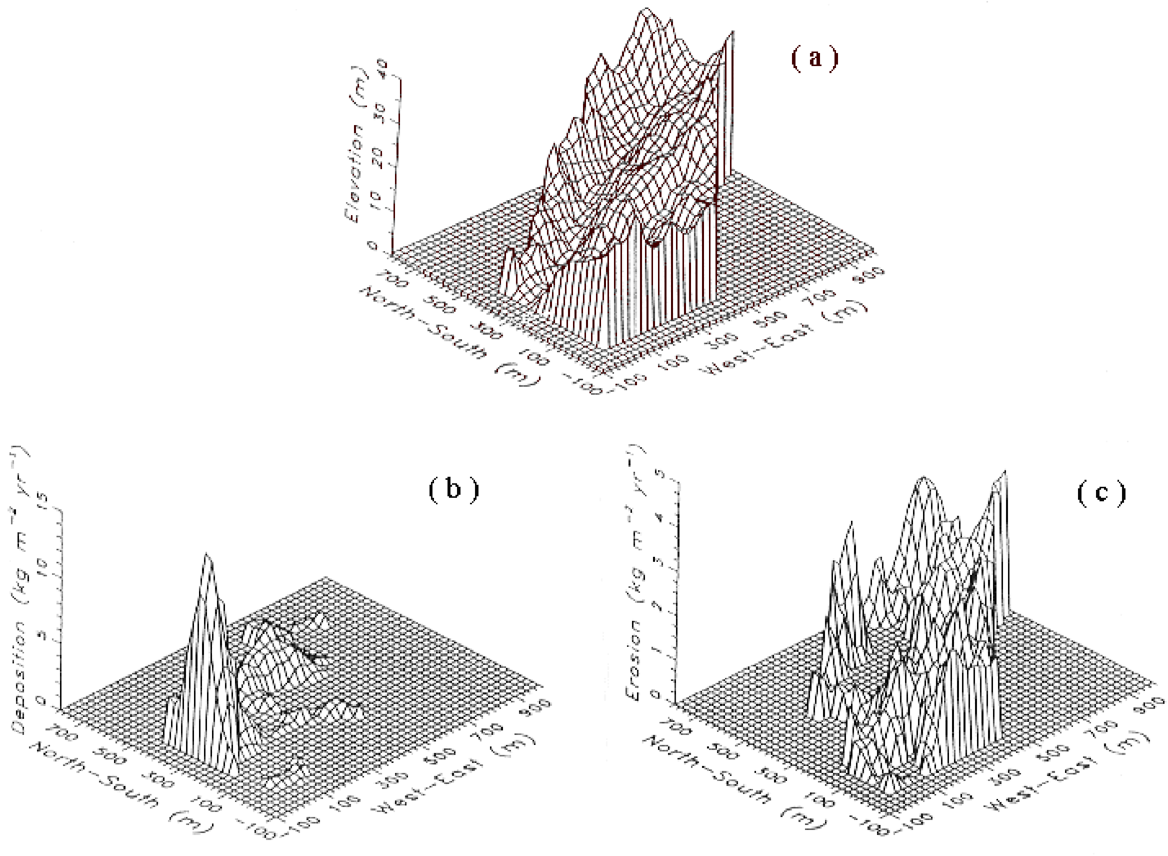


Figure 1 (a) Relative elevation of Palouse study site (15x vertical exaggeration), (b) soil deposition rates, and (c) soil erosion rates across study area based on kriged estimates (from Busacca et al. 1993, 365-366).

The final two assumptions are also important because they affect how the GIS divides the landscape into zones and how the attributes (model inputs) are aggregated (estimated) in each zone. The original USLE computed average soil loss along hillslope profiles that were defined with reference to a "standard" soil loss plot. These standard plots were 22.1 m long and planar in form although these conditions may not occur very often in natural landscapes (Moore and Wilson 1994). Foster and Wischmeier (1974) later divided irregular slopes into a series of uniform segments and modified the original USLE *LS* equations to calculate the average soil loss on these slope profiles. However, this method still requires the subdivision of landscapes into hillslope facets. Griffin et al. (1988) rewrote the original USLE to calculate erosion at any point in a landscape and thereby avoided this requirement. Their equation is much easier to implement than the original model, although the user must still distinguish those areas experiencing net erosion and deposition. This version also retains the 1-D structure of the original model and (similar to the original USLE and RUSLE models) cannot handle variations in runoff rates caused by spatially varying infiltration rates (Kinnell et al. 1995) and/or converging and diverging terrain (Moore and Wilson 1992, 1994, Wilson and Gallant 1996).

None of the GIS-based USLE applications discussed above mentioned the need to distinguish areas experiencing net erosion and deposition before applying this equation. It is not clear how they responded to this challenge (if at all) and the discussion about scale (i.e. size of raster cells and/or vector polygons used to compute soil loss) and the consequences of using source data compiled at different scales is also vague. These applications also used different slope gradient and length terms from those specified in the original model, assumed that runoff was generated uniformly across the landscape, and ignored the revised USLE proposed by Griffin et al. (1988) for estimating soil erosion at points (grid cells) in the landscape (Wilson 1996).

Repetto and Wilson (1996) calculated five sets of RUSLE *LS* values for a 2,900 km² catchment in southwest Montana and found that the magnitude and spatial pattern of LS varied greatly with data source (30 m and 3 arc-second USGS DEMs) and grid spacing (30, 100, and 200 m grid cells). The five DEMs used for this study were prepared with ANUDEM (Hutchinson 1989). This program automatically removes spurious pits within user-defined tolerances, calculates stream lines and ridge lines from points of locally maximum curvature on contour lines and (most importantly) incorporates a drainage enforcement algorithm to maintain fidelity with a catchment's drainage network (Hutchinson 1989, Moore et al. 1993b). *LS* values were grouped into five categories and the overall agreement between pairs of maps was defined as the percentage of cells that were assigned to the same *LS* classes. Table 1 lists the overall agreement values for different pairs of maps and shows that low levels of overall agreement (< 50%) were computed for each pair of maps.

Table 1. Overall agreement between RUSLE *LS* values at five mapping resolutions (%) (from Repetto and Wilson, 1996)

	Data source & map resolution				
Data source & map resolution	30m/ 30m	30m/ 100m	3 a-s/ 100m	30m/ 200m	3 a-s/ 200m
30 m DEM / 30 m spacing	---				
30 m DEM / 100 m spacing	31.3	---			
3 arc-second DEM / 100 m spacing	27.4	37.5	---		
30 m DEM / 200 m spacing	27.2	40.1	37.4	---	
3 arc-second DEM / 200 m spacing	26.9	38.7	42.6	49.3	---

ANSWERS

The Areal Nonpoint Source Watershed Environmental Response Simulation (ANSWERS) model was developed by Beasley and Huggins (1982) to simulate surface runoff and erosion in predominantly agricultural catchments. The model divides catchments into square elements (grid cells) and uses the connectivity of the cells (derived from slope aspect values) and the continuity equation to route flow to the catchment outlet (Beasley et al. 1982). Three erosion processes are considered: detachment of soil particles by raindrop impact, detachment of soil particles by overland flow, and transport of soil particles by overland flow. The quantity of erosion or deposition occurring within each cell is estimated based on the erodibility of the soil and land cover type of the cell, the rate of flow passing through the cell, and the quantity of sediment in the flow passing through the cell (Brown et al. 1993). A series of topographic (elevation, slope, aspect), soil (porosity, moisture content, field capacity, infiltration capacity, USLE *K* factor), land cover (percent cover, interception, USLE *CP* factor, surface roughness, retention), channel (width, roughness), and rainfall inputs are required for each element (De Roo et al. 1989).

The original version of ANSWERS was limited to 20 spatially homogeneous soil and land cover types because the input files had to be created by hand (Beasley and Huggins 1982, Beasley et al. 1982). The collection and organization of the input data in a GIS means that unique values can be used for each element and the level of spatial aggregation is determined by the size of the grid cells (De Roo et al. 1989, Joao and Walsh 1992). This fundamental change in data resolution and model organization will almost certainly alter the output of the model: De Roo et al. (1989), for example, found that ANSWERS predicted 46% more total runoff and 36% more soil loss when a GIS-based version of the model that divided the landscape into 4,275 0.01 ha (10 m by 10 m) grid cells was used in place of the original (lumped) model for a measured rainfall event in the Catsop catchment located in Limburg Province, Netherlands.

Brown et al. (1993) also examined the response of this model to variations in input data aggregation levels for a 2,100 ha catchment in the central piedmont of North Carolina. Five GIS data layers representing land cover, soil type, slope angle, slope aspect, and stream channels were prepared and used with a series of look up tables to derive model inputs for 23,629 0.1 ha (30 m by 30 m) grid cells (Joao and Walsh 1992). Soils, land cover, and slope angle coverages were then generalized by assigning the class value occupying the majority of the area within an aggregation unit to all cells within that unit at eight generalization levels (see Table 2 for details).

Table 2. Overall agreement between erosion and deposition at eight mapping resolutions (%) (from Brown et al. 1993, 507)

	Generalization level							
Generalization level	30	60	120	180	240	300	420	600
30	---							
60	84.8	---						
120	75.9	72.2	---					
180	65.7	66.5	67.7	---				
240	62.8	63.0	64.9	69.8	---			
300	61.2	61.3	62.6	66.2	69.7	---		
420	55.9	56.4	58.5	64.2	67.6	64.1	---	
600	53.8	54.7	56.5	60.0	64.4	62.6	69.3	---

Slope aspect and stream channel coverages were not generalized to maintain the connectivity of the hydrologic network. The patterns of semivariogram and fractal dimension plots were similar among soil and land cover parameters and indicated that the surface variation was more dependent on the pattern of polygon boundaries captured with the original data sources than on the actual attribute values represented by those polygons (land cover data were interpreted and digitized from 1:58,000-scale color aerial photographs and soils data were digitized from 1:15,480-scale NRCS maps for this study).

Brown et al. (1993) also implemented the ANSWERS model with each set of input data and a user-defined precipitation event to produce a series of erosion/deposition maps. Erosion and deposition were grouped into four categories and the overall agreement between pairs of maps was defined as the percentage of cells that were assigned to the same erosion/deposition classes. Table 2 lists overall agreement values for different pairs of maps and shows: (1) the best overall agreement occurred between model runs with low aggregation levels (30-60 m cell spacings); (2) overall agreement between maps decreased as resolution differences increased; (3) overall agreement decreased with increasing cell size (the 30 and 600 m runs were least similar for example). Percent area curves for erosion and deposition in the whole basin and along a stream corridor (reproduced in Figure 2) show that the break in the erosion and deposition estimates consistently occurred between the model runs with 120 m by 120 m and 180 m by 180 m input cells. This range of cell sizes corresponded to the sampling interval at which spatial dependance was maintained and would seem to indicate that smaller cell sizes (30-60 m on a side) were unnecessarily detailed given the input data. The effects of aggregation on model output was relatively minor to the 120 m by 120 m cell size and any cell size within this range (30 by 30-120 by 120 m) was likely to produce similar model results.

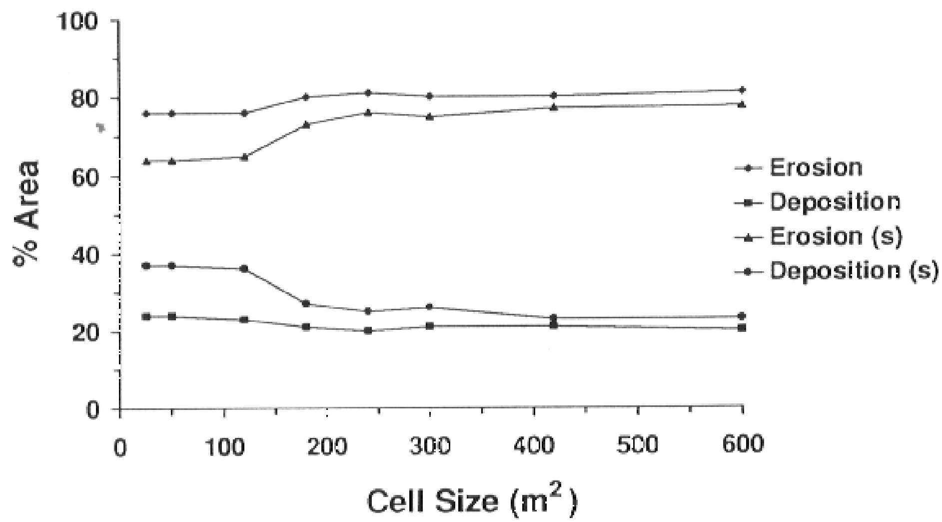


Figure 2. Percent area of erosion and deposition as affected by data aggregation. Curves are given for the basin as a whole and for all areas within 60 m of stream channels [indicated by (s)] (from Brown et al. 1993, 507).

De Roo et al. (1989) also conducted multiple ANSWERS model runs to evaluate the sensitivity of surface runoff and soil loss predictions to individual soil, land cover, and channel inputs. Topographic inputs were computed from a DEM and the other factor values were estimated from field (point) observations and/or geostatistical interpolation techniques (e.g. block kriging) at a 10 m grid spacing. One model run used average soil and land cover inputs to simulate a 20 minute, 20 mm storm in a 42.7 ha portion of the Catsop catchment (Table 3). Five values above and below the average values were chosen to represent possible values for soils and vegetation in the region and used in ten additional model runs. A simple sensitivity index was computed to describe the output variation $(RES_{11} - RES_1)$ around an average output (RES_6) as follows:

$$S_{(V1,V11)} = (RES_{11} - RES_1) / \mathrm{RES}_6 \qquad (2)$$

where RES_1 and RES_{11} are the results produced with the smallest and largest values used for each of the input variables. Table 3 shows that the model is very sensitive to the infiltration variables and antecedent soil moisture content. This is a potentially serious problem because: (1) these variables can be expected to vary through space and time; (2) measurement and interpolation of these variables is difficult and expensive; and (3) results from individual studies may be difficult to extrapolate to other rainstorms and/or catchments because the relationships between model inputs and outputs are non-linear and cannot be described by one type of function (Figure 3). The sensitivity of model outputs to these input variables can be expected to vary with the precipitation patterns, different rainfall-intensity distributions, and topography (relief, landforms, etc.) within the simulated catchment.

Table 3. Results of a sensitivity analysis of the ANSWERS model (from DeRoo et al. 1989, 256)

Variable	Range	Average Value	Sensitivity index Runoff	Soil Loss
TP (%)	40-50	45	-1.89	-1.79
FP (%)	50-70	60	0.24	0.26
FC (mm/h)	2-18	10	-3.98	-5.88

A (mm/h)	20-300	160	-10.83	-26.23
DF (mm)	5-155	80	-11.37	-25.21
P (--)	0.6-0.7	0.65	0.39	0.67
ASM (%)	40-90	65	10.15	22.32
K (--)	0.4-0.7	0.55	0	0.44
PIT (mm)	0-3	1.5	-1.67	-2.18
PER (%)	0-100	50	-1.05	-1.56
RC (--)	0.3-0.8	0.55	-2.11	-4.81
HU (mm)	30-180	105	-0.49	-1.19
N (--)	0.05-0.37	0.21	-2.62	-4.02
C (--)	0.1	0.5	0	1.79
WIDTH (m)	0.5-45	2.5	-0.06	-0.77
MAN. (--)	0.01-0.16	0.085	-0.09	-0.07
GRF (--)	0-0.001	0.0005	0.94	0.11

TP total porosity
FP field capacity
FC saturated infiltration capacity
A initial minus saturated infiltration capacity
DF infiltration control-zone depth
P infiltration constant
ASM antecedent soil moisture
K soil erodobility (USLE K)
PIT potential interception capacity
PER crop coverage
RC soil roughness coefficient
HU max roughness height
N Manning's n soil surface
C crop factor (USLE C*P)
WIDTH channel width
MAN Manning's n channel
GRF groundwater release fraction

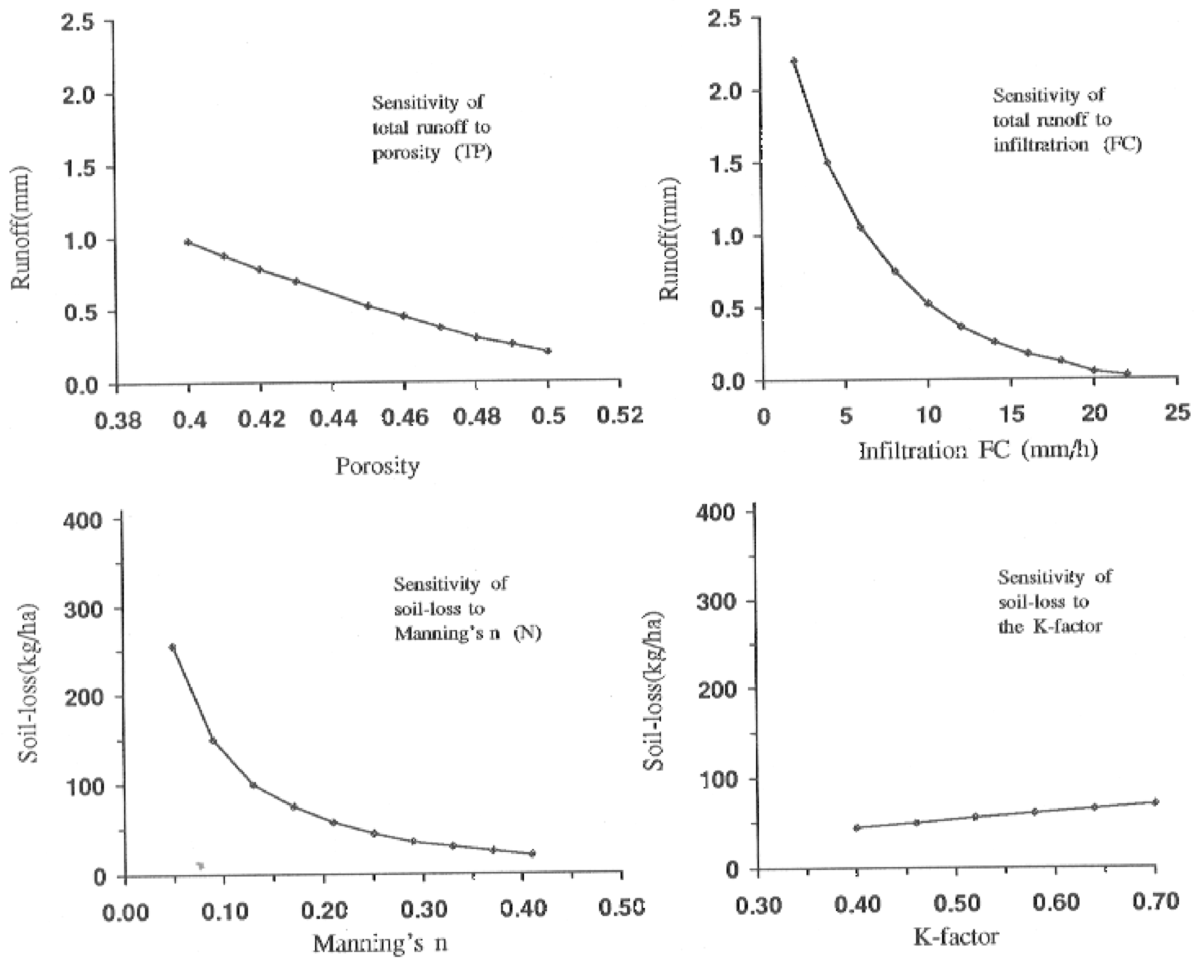


Figure 3. Sensitivity of selected ANSWERS model ouputs to input variables (from De Roo et al. 1989, 525)

AGNPS

The AGricultural NonPoint Source (AGNPS) model is an event-based model that simulates runoff, sediment, and nutrient transport from agricultural watersheds ranging in size from a few hectares to approximately 20,000 ha (Young et al. 1989). The model incorporates separate hydrology, erosion, sediment transport, and chemical transport modules which route water, sediments, and other contaminants through cells from the catchment boundary to the outlet in a stepwise fashion. The choice of cell size affects the time and labor required to run the model as well as model accuracy, and 16 ha (400 m by 400 m) grid cells are recommended for watersheds exceeding 800 ha. Runoff volumes are estimated with the USDA-NRCS curve number method and upland erosion is estimated with a modified version of the USLE (Young et al. 1987). Sediment transport and deposition are estimated with equations proposed by Foster et al. (1981) and Lane (1982) respectively, and the chemical transport component is based on relationships derived for CREAMS (Frere et al. 1980) and a feedlot evaluation model (Young et al. 1982). Several erosion (streambanks, streambeds, gullies, etc.) and nutrient sources (animal feedlots) are treated as point sources and added to contributions from diffuse sources. Model outputs can be obtained for each cell and/or at the watershed outlet.

Engel et al. (1993) linked AGNPS with the GRASS GIS as part of a decision support system to assist with the management of runoff, erosion, and nutrient movement in agricultural landscapes. The GIS was used to organize the input data and display the model results in this application. Terrain-based attributes (slope gradient, length, shape, aspect, upslope contributing area, etc.) represent approximately one-third of the input parameters required by this model and can be generated from DEMs. Panuska et al. (1991) used a series of contour- and grid-based terrain analysis methods and DEMs to generate a terrain-based parameter file that could be linked with AGNPS. They also examined the sensitivity of the slope, upslope contributing area, and maximum flow path length variables computed from different DEM structures for a range of element sizes on the 210 ha North Fork Cottonwood Creek catchment in southwestern Montana. The computed flow path lengths and contributing areas (but not slopes) varied with terrain analysis method and element (cell) size (Figure 4). Panuska et al. (1991) concluded that the adequacy of the DEM would depend on the characteristics of the terrain since the smallest cell size will only produce the most accurate representation of the terrain if the dimensions of the cells are greater than the horizontal resolution of the primary elevation data and the elevation differences between neighboring points are greater than the vertical resolution of the data. This result suggests that rolling terrain with small to moderate relief would be better represented by a coarser DEM than a dissected catchment with steep ridges and ravines.

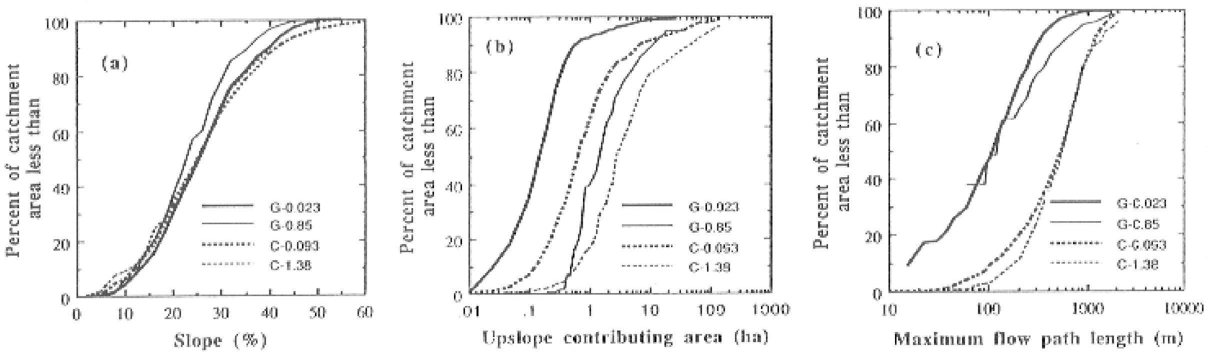


Figure 4. Examples of (a) slope, (b) upslope contributing area, and (c) maximum flow path length distributions calculated with (G) grid- and (C) contour-based DEMs and terrain analysis methods for a range of element sizes (from Panuska et al. 1991)

Garbrecht and Martz (1994) also used sensitivity analysis to examine the dependence of six drainage properties (critical source area, number of channel links, total channel length, mean channel link slope, watershed drainage density, and mean channel link direct drainage area) on DEM cell size. They started with a 30 m DEM in central Oklahoma and generated coarser DEMs by successive spatial averaging of the baseline DEM. Their results showed that a DEM would need to have a grid area of less than 5% of the network reference area (i.e. the mean area draining directly into the channel links) to reproduce important drainage features within 10% of the baseline reference values because the DEM cannot accurately reproduce drainage features that are at the same scale as the spatial resolution of the DEM. These results are similar to those of Panuska et al. (1991) to the extent they show that: (1) the grid size must be selected relative to the size of the features of interest, and (2) high-resolution DEMs are needed if small drainage features are important.

CMLS

The Chemical Movement through Layered Soils (CMLS) model was developed by Nofziger and Hornsby (1986, 1987) to interactively simulate chemical movement through soil with easily obtained soil, chemical, and weather inputs. CMLS divides the soil into as many as 20 layers and calculates the fraction of the applied chemical remaining in the entire soil profile and the position of the solute front at different times based on the piston displacement of water. The soil properties affecting chemical movement (bulk density, permanent wilting, field capacity water contents, and organic carbon content) may vary between layers but are assumed to be uniform within each layer. Two chemical properties (the partition coefficient normalized to soil-organic carbon and degradation half-life) and several climatic and cultural factors known to affect chemical movement (plant root depth, daily precipitation, irrigation, and evapotranspiration amounts) are also required by the model. Although this model was written primarily as a management and educational tool, its performance has been compared favorably with observed data and the predictions of several other pesticide fate models in several U.S. locations (Pennell et al.

1990, Inskeep et al. 1996). The CMLS model has also been combined with GIS to predict the threat to groundwater posed by current herbicide applications (e.g. Wilson et al. 1993) and several recent studies have examined the impact of data resolution and input data estimation methods on model outcomes. The problems are complicated in the U.S. because modern soil surveys report large ranges for most soil properties.

Foussereau et al. (1993) used bootstrapping to generate a series of pseudo-profiles of soils from pedon characterization data and evaluate the uncertainty of CMLS model predictions due to variability of soil input data beneath citrus groves in southwestern Florida. Their method is important because it shows how the variability of the major soils occurring in map units can be incorporated in groundwater pollution assessments. Five hundred pseudo-soil profiles were generated for each single-name soil map unit (known as consociations) from three or more actual pedon characterization data sets and combined with weather sequences predicted with the WGEN weather generator (Richardson and Wright 1984) to produce cumulative probability curves showing the fraction of applied pesticide leaching beyond the 1 m depth (Figure 5). At least three actual pedon characterization data sets were also used to generate pseudo-profiles for each of the named soils in multiple-named soil map units (i.e. soil associations or soil complexes) and the highest probability of exceedance (POE) of the U.S. EPA health advisory levels for specific pesticides was used as an estimator of the environmental risk on GIS maps showing these map units. Other soil inclusions that were not included in map unit names (although they may represent $\leq$25% of a single-name soil map unit) were not considered because map unit descriptions in soil survey reports usually contain only qualitative descriptions of these soils. Foussereau et al. (1993) did compile some transect data for three Florida map units that showed how a given map unit may have different types and percentages of soils from those provided in the soil map unit descriptions in soil survey reports. These discrepancies left them with continuing and difficult questions as to how they might incorporate these soils into their bootstrapping algorithm without creating hypothetical soil profiles that do not reflect the actual soils occurring in the landscape of interest.

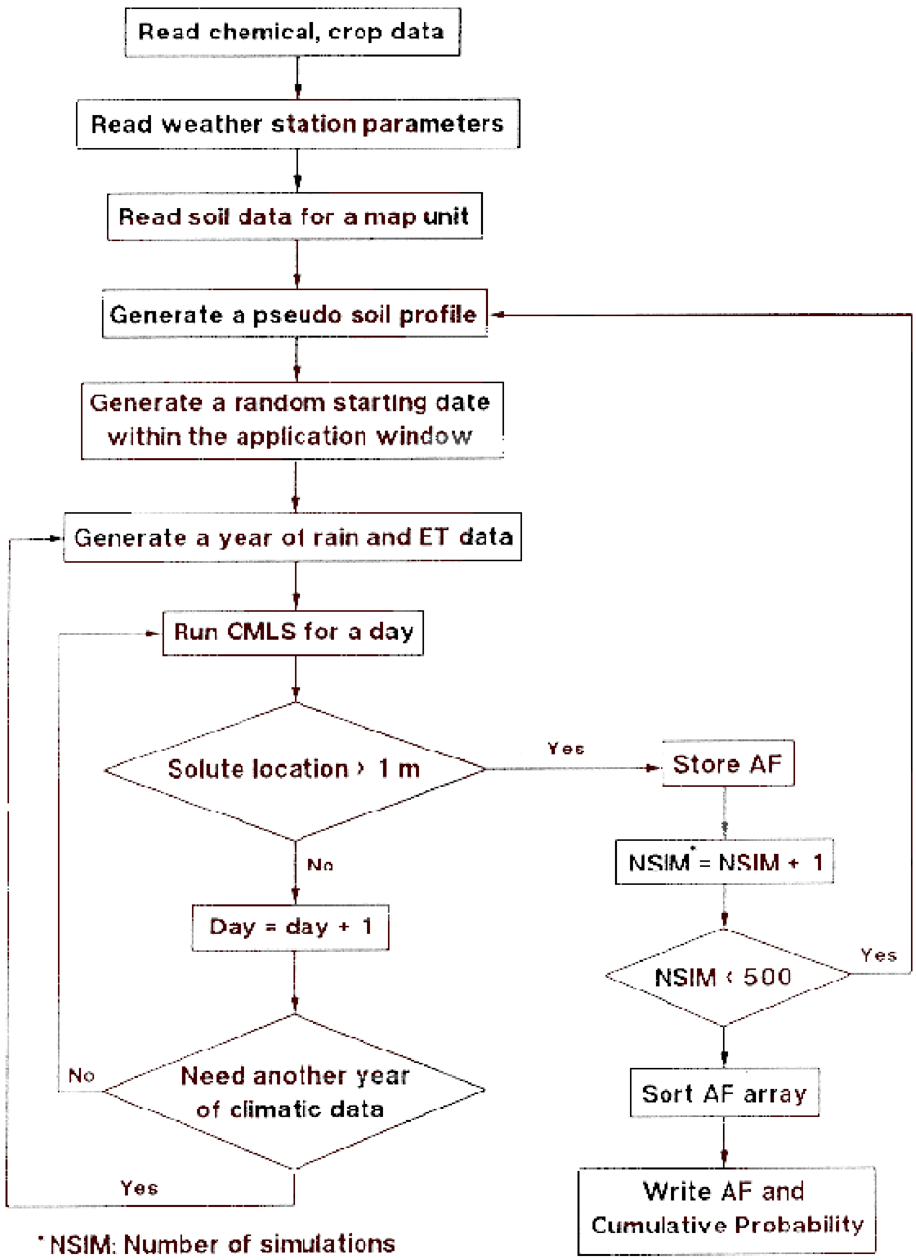


Figure 5. Flow chart of simulation procedure using pseudo-profile data, generated weather sequences, estimated evapotranspiration, and varying pesticide appllications (from Foussereau et al. 1993, 265).

Wilson et al. (1996) used the WGEN and CMLS models with two sets of soil and climate inputs to evaluate the impact of input data resolution on model predictions for a 320 km² study area in Teton County, Montana. The basic soil and climate inputs required by WGEN and CMLS were acquired from either: (1) the USDA-NRCS State Soil Geographic (STATSGO) database (Bliss and Reybold 1989); (2) the USDA-NRCS (County) Soil Survey Geographic (SSURGO) database (Reybold and TeSelle 1989); (3) the Montana Agricultural Potentials System (MAPS) database (which divides Montana into approximately 18,000 20 km² cells and stores more than 200 different land and climate characteristics for each of these cells) (Nielsen et al. 1990); and (4) a series of fine-scale monthly climate surfaces developed by the authors (0.55 km² cell size) using thin-plate splines, published climate station records and USGS DEMs (Custer et al. 1996) (Figure 6). Fifteen years of daily precipitation and evapotranspiration values were generated and combined with soil and pesticide inputs in CMLS to estimate the depth of picloram (4-amino- 3,5,6-trichloro-2-pyridinecarboxylic acid) movement at the end of the growing season for every polygon containing unique soil and/or climate inputs. The results showed that: (1) the mean depths of picloram movement predicted for the study area with the SSURGO soil and MAPS (coarse-scale) climate information and the two model runs using the fine-scale climate data were significantly different from the values predicted with the STATSGO soil and MAPS climate data (based on a new variable representing the differences between the depths of leaching predicted with the different input data by soil/climate map unit and testing whether the mean difference was significantly different from zero at the 0.01 significance level) (Table 4); and (2) CMLS identified numerous (small) areas where the mean center of the picloram solute front was likely to leach beyond the root zone when the county soils information was used (Figure 7). These results show how the CMLS model predictions vary with the choice of climate and soil inputs and why high resolution SSURGO soil information is needed if the goal is to identify those areas where potential chemical applications are likely to contaminate groundwater.

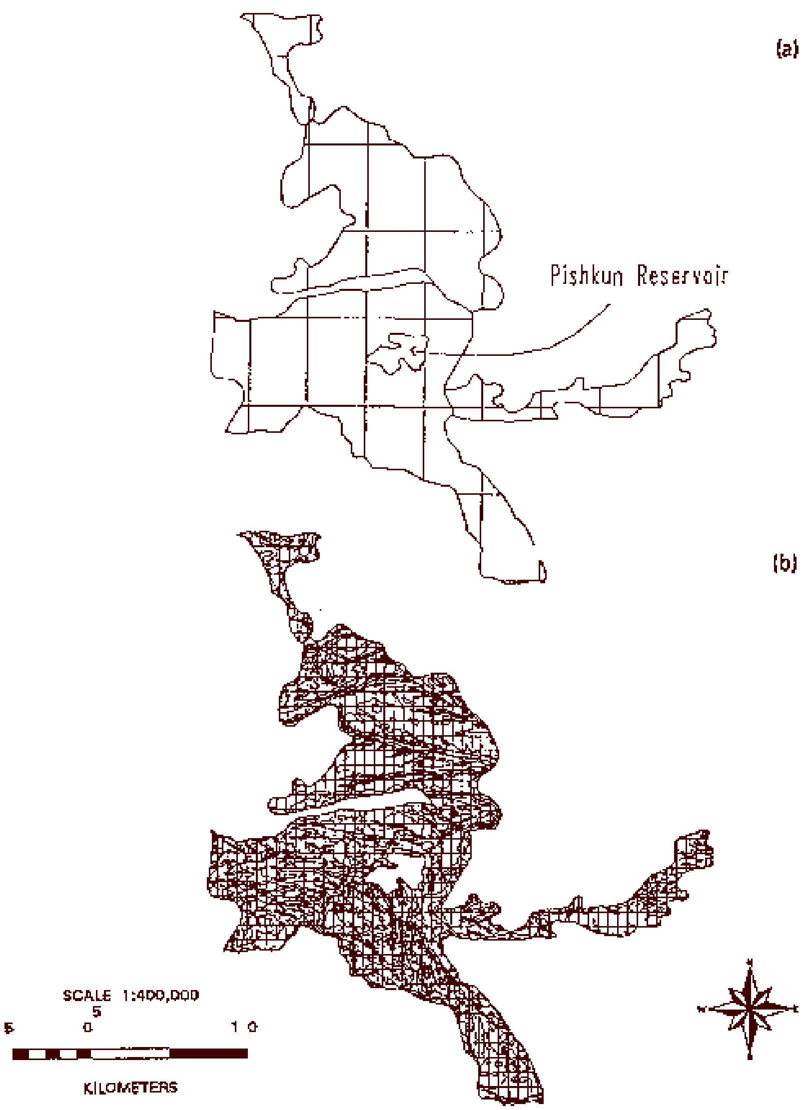


Figure 6. Maps of Teton County study area showing (a) MAPS climate cells overlaid on STATSGO soil mapping units, and (b) ANUSPLIN climate cells overlaid on SSURGO soil mapping units (from Wilson et al. 1996).

Table 4. T test results comparing MAPS/SSURGO, ANUSPLIN/STATSGO, and ANUSPLIN/SSURGO CMLS model prediction with MAPS/STATSGO CMLS model predictions (from Wilson et al. 1996).

Model run	Mean depth (cm)	Std. error	T	Prob > |T|
MAPS/SSURGO	2.52	0.43	5.80	0.0001
ANUSPLIN/STATSGO	-1.37	0.19	-7.12	0.0001
ANUSPLIN/SSURGO	1.65	0.47	3.49	0.0010

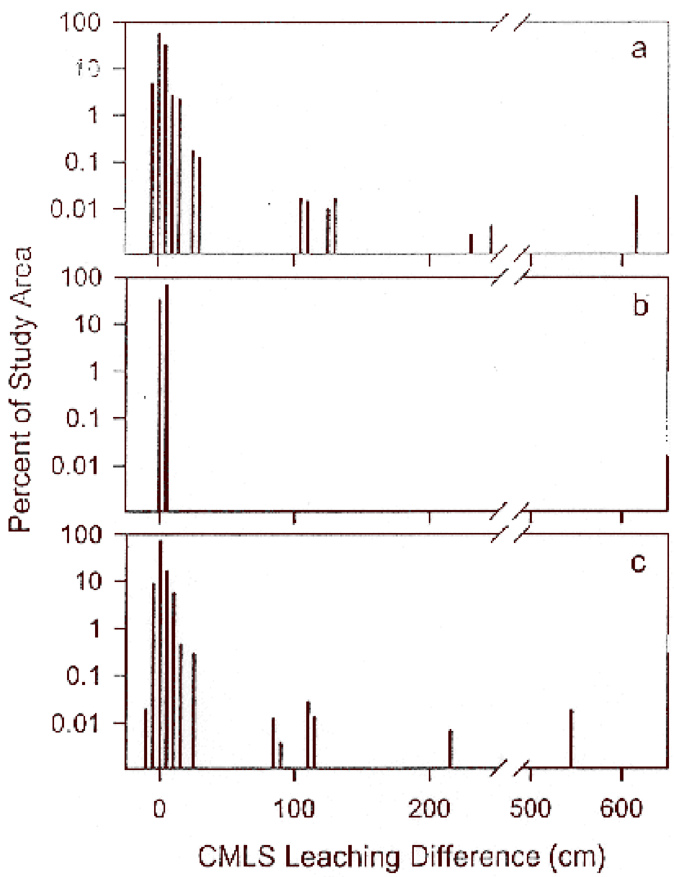


Figure 7. Histograms showing differences in predicted depth of solute movement for MAPS/STATSGO and (a) MAPS/SSURGO, (b) ANUSPLIN/STATSGO, and (c) ANUSPLIN/SSURGO data inputs (from Wilson et al. 1996).

Inskeep et al. (1996) compared predicted and observed pentafluorobenzoic acid (PFBA), 2, 6-difluorobenzoic acid (DFBA) and dicamba travel times at a single field site near Manhattan, Montana. CMLS and LEACHM (Wagenet and Hutson, 1989) predictions were generated using: (1) detailed site-specific measurements (both models); (2) conductivity and retentivity functions estimated from the SSURGO database (LEACHM model); and (3) volumetric water contents estimated from textural data in the SSURGO database and daily precipitation and evapotranspiration estimated with the WGEN weather generator and MAPS database (CMLS model). Comparison of observed and simulated mean travel times showed that: (1) both LEACHM and CMLS performed adequately with site-specific inputs (e.g. CMLS predicted mean travel times were within 3.5 to 38% of observed data over two growing seasons under a variety of crop and fallow conditions), and (2) the CMLS predictions were less sensitive to data input resolution than the LEACHM predictions due in part to the fact that CMLS provides an over-simplified description of transport processes. Inskeep et al. (1996) concluded that the use of the SSURGO and MAPS databases with CMLS may provide a reasonable approach for classifying the susceptibility of map units in terms of solute movement, although the potential applicability of this approach is limited to areas with digital copies of the SSURGO (selected counties scattered throughout the U.S.) and MAPS databases (Montana only).

LEACHM

The Leaching Estimation and CHemistry model (LEACHM) is a one-dimensional finite difference model designed to simulate the movement of water and solutes through layered and non-layered soil profiles (Wagenet and Hutson 1989). This is deterministic, mechanistic, research-oriented model with correspondingly greater input requirements than many simpler models (Inskeep et al. 1996). The model uses: (1) a variable time step based on allowable water content changes in the soil profile; (2) Darcy's law and the continuity equation to describe transient water flow; and (3) calculated water contents and fluxes to solve the convection-dispersion equation and describe the movement of solutes which can adsorb, volatilize, and degrade. The model also allows depth- and time-dependent root growth, water use (transpiration), and evaporation (Wagenet and Hutson 1989). LEACHM has been validated and used as a predictive tool at the plot and field scale (e.g. Wagenet and Hutson 1986, Wagenet et al. 1989), and several attempts have been made to combine this model with GIS databases for regional scale assessments of leaching behavior.

Petach et al. (1991) used LEACHM to simulate the movement of four chemicals through layered soils for a 70 km^2 study area near Albany, New York. The site was divided into 0.4 ha grid cells and a series of pedotransfer functions was used to relate soil physical properties to the mean and variance of the hydraulic properties occurring in each cell. The soils were classified into six hydraulic groups and LEACHM was executed 25 times using different combinations of soil hydraulic properties to represent the expected spatial variability in hydraulic properties. This approach generated estimates of the variability of chemical fluxes near the low end of the range published in the literature (although the simulated variability could well agree with field measured values if the process was extended to include other factors that affect chemical movement such as sorption and degradation) and the influence of the soil hydraulic properties on the estimated variability in leaching of both water and chemicals was approximately the same magnitude as the impact of the two precipitation years considered in this study.

The experience gained in the above study led to the development and application of a modified version of the pesticide version of LEACHM to a 300,000 km^2 study area encompassing the states of Connecticut, Maine, Massachusetts, New Hampshire, New York, Rhode Island, and Vermont (Hutson and Wagenet 1993). This new model (LEACHA) replaced the Richards equation for water flow and the convection-dispersion equation for chemical transport with a mobile-immobile capacity model adapted from Addiscott (1977) and Nicholls et al. (1982). LEACHA divides the soil profile into horizontal layers and uses a daily time step to calculate fluxes and changes in water and chemicals for each layer. Soil attributes (mean clay content and bulk density values and the lowest (worst-case) organic carbon content) were obtained from the STATSGO database and water retention properties were estimated using regression equations developed by Rawls and Brakensiek (1982) which relate water retention to particle size, bulk density and organic matter data. The results in the New England application were mapped using four classes of leaching potential and showed that 60-84% of the variability in leaching could be explained by differences in soil organic carbon content (Hutson 1993). The other 16-40% of the variability was ascribed to differences in climate and parent material (Wagenet and Hutson 1996).

Overall, these results indicate why additional knowledge and documentation of precipitation and soil organic matter content patterns may be required to improve regional scale applications of LEACHA and other solute transport models. The precipitation inputs used by Hutson (1993) could have been improved using one of the new interpolation methods that predicts spatial patterns of precipitation based on published station records and DEMs (i.e. latitude, longitude, and elevation data) (e.g. Daly et al. 1994, Hutchinson 1995, Running and Thornton 1996). The soil organic carbon variable represents a more difficult problem because the soil organic matter data reported in published soil surveys are very generalized and may not reflect the soil organic matter conditions for the landscape unit being modelled (Wagenet and Hutson 1996).

TOPMODEL

TOPMODEL predicts the relative amount and spatial distribution of subsurface, infiltration excess, and saturation excess overland flow based on surface topography and soil properties (Beven and Kirkby 1979, Beven et al. 1984, Sivapalan et al. 1987, Quinn and Beven 1993). The model has been validated with rainfall-discharge data (e.g. Beven et al. 1984, Hornberger et al. 1985, Robson et al. 1993, Obled et al. 1994, Wolock 1995) and several recent studies have examined its applicability to water quality problems (Wolock et al. 1990, Robson et al. 1992). The continued popularity of TOPMODEL can be traced to its structural simplicity and parsimonious parameterization (Iorgulescu and Jordan 1994).

The model assumes a spatially uniform recharge rate and quasi-steady subsurface response to derive a function relating local soil moisture storage or water table depth to the topographic index (ln(*a*/tanß)) of a catchment:

$$S_i = S + m\{\lambda - \ln(a/\tan\beta)_i - (\delta - \ln(K_i)\} \quad (3)$$

where S_i is the local soil moisture deficit, S is the mean soil moisture deficit of the catchment, m is a parameter that characterizes the decrease in hydraulic conductivity with soil depth, a is the drainage area per unit contour length, ß is the slope, K_i is the lateral transmissivity of the soil profile when the water table just intersects the surface, and λ and δ are the mean values of ln(*a*/tanß) and ln(*K*) for the catchment (Zhang and Montgomery 1994). Many applications ignore the soil transmissivity terms in (3) because the spatial pattern of soil transmissivity is seldom known and is often assumed to be constant over the catchment (Iorgulescu and Jordan 1994). S_i represents a negative soil moisture deficit so that $S_i = 0$ at complete saturation and $S_i > 0$ when a soil moisture deficit occurs. The mean soil moisture deficit of a catchment at time t, S_t, is calculated by:

$$S_t = S_{t-1} - (q_{t-1} - r)\Delta t \quad (4)$$

where q is the total catchment runoff at time t - 1 divided by the catchment area, r is the net recharge rate into the soil column, and Δt is the time interval used for the model simulation. The soil moisture deficit at every point (grid cell) in the catchment is then computed using (3) and water is routed to the catchment outlet via: (1) subsurface runoff in areas with a soil moisture deficit larger than the precipitation added during a time step; (2) subsurface and infiltration excess overland flow in areas with rainfall intensities greater than the infiltration capacity; and (3) subsurface and saturation excess overland flow in areas with either a soil moisture deficit smaller than the incremental precipitation in a unit time step or that were saturated during the previous time step (Beven et al. 1984). The subsurface flow rate q_b of the catchment is calculated by:

$$q_b = \exp(-(\lambda - \delta))\exp(-S_t/m) \quad (5)$$

and the saturation excess runoff q_0, which is the sum of the excess soil moisture and direct precipitation that falls on the saturated areas, is calculated by:

$$q_0 = (1/A_t)\int_{As}\{-S_i/\Delta t + r\}\,dA \quad (6)$$

where A_s is the area of the catchment with surface saturation ($S_i \le 0$) and A_t is the total area of the catchment. This approach means that predicted soil moisture patterns will follow the outline of the topographic index and the predicted saturated source area will expand and contract as the water balance of the model changes (Quinn et al. 1995). Total runoff q at each time step is the sum of subsurface and surface runoff (Beven and Kirkby 1979, Zhang and Montgomery 1994).

The ln(*a*/tanß) index was calculated manually using contour data in early applications of TOPMODEL and the advent of GIS and terrain analysis techniques has allowed this procedure to be automated (Quinn et al. 1995). However, several recent studies have demonstrated that the spatial pattern and statistical distribution of the topographic index varies with different grid resolutions and estimation procedures.

Zhang and Montgomery (1994) calculated slope, drainage area per unit contour length, and the topographic index for a series of depressionless square-grid DEMs at scales of 2, 4, 10, 30, and 90 m using the GRID tools in ARC/INFO and spot elevation data obtained from low-altitude aerial photographs. They found that the DEM grid size significantly affected both the computed topographic parameters and hydrographs for two study areas with moderate to steep relief in the western United States. The 10 m grid spacing provided a substantial improvement over the 30 m and 90 m data, but the 2 m and 4 m data provided only marginal additional improvement. This last result may be a function of the scale of the source data (spot elevations derived from low altitude aerial photography with a stereo digitizer at a density of approximately 10 m) and/or the method used to calculate drainage areas (i.e. the classical D8 algorithm discussed below). Moore (1996) showed that computed slope and topographic index values varied with grid size for a series of 22 square-grid DEMs with scales from 20 to 680 m in three 100 km² study areas in southeastern Australia. Quinn et al. (1995) computed drainage areas for a series of 5, 10, 25, and 50 m DEMs derived from a 1:10,000-scale contour map and found that: (1) small channels and catchment boundaries tend to become obscured or lost altogether as grid size increases, and (2) larger grid sizes exhibit a bias towards larger topographic index values. Zhang and Montgomery (1994) and Quinn et al. (1995) both concluded that DEMs of the order of 10 m or smaller were needed to capture the variability of the topographic form for hillslopes in their study areas, and the continued evolution and spread of modern Global Positioning System (GPS) tools may allow the collection of high resolution topographic data sets in some environments (Spangrud et al. 1995).

The choice of flow routing method may also affect the magnitude and spatial pattern of the computed topographic index values, and some programs provide multiple options. The TAPES-G user, for example, can select either the classical D8 algorithm (O'Callaghan and Mark 1984), the quasi-random Rho8 algorithm (Fairfield and Leymarie 1991), the multiple flow direction FD8/FRho8 algorithm (Freeman 1991, Quinn et al. 1991), or the stream-tube based DEMON algorithm of Costa-Cabral and Burges (1994) for calculating upslope contributing areas (Gallant and Wilson 1996).

The D8 algorithm allows flow from a node to only one of eight nearest neighbors based on the direction of steepest descent. The D8 algorithm tends to produce flow in parallel lines along preferred directions, which will only agree with the aspect when aspect is a multiple of 45°, and it cannot model flow dispersion (Moore et al. 1993a). However, this algorithm is used in ARC/INFO and remains the most commonly used algorithm notwithstanding these limitations. The Rho8 algorithm developed by Fairfield and Leymarie (1991) is a stochastic version of the D8 algorithm in which the expected value of the flow direction is equal to the aspect. This algorithm cannot model flow divergence (like the D8 algorithm), but it does simulate more realistic flow networks because long parallel flow paths are broken up at the cost of many more cells without upslope connection (Moore et al. 1993a).

The FD8 and FRho8 algorithms allow flow divergence or catchment spreading to be represented in upland areas above defined channels and use the D8 or Rho8 algorithms below points of channel initiation. The proportion of flow or upslope contributing area assigned to multiple downslope nearest neighbors above channels is determined on a slope-weighted basis using methods similar to those proposed by Freeman (1991) and Quinn et al. (1991). The FRho8 option gives much more realistic (smoother) distributions of contributing area and also eliminates D8's parallel flow paths (Quinn et al. 1991, Moore et al. 1993a, Wolock and McCabe 1994, Gallant and Wilson 1996). Because stream lines are usually quite well defined, the flow divergence (spreading) algorithm is usually best disabled in areas of high contributing area, and TAPES-G provides a threshold called the "maximum cross grading area" above which the simple D8 or Rho8 method is applied (Gallant and Wilson, 1996).

The DEMON algorithm generates flow in each pixel (source pixel) and follows it down a stream tube until the edge of the DEM or a pit is encountered. These stream tubes expand and contract in response to the DEM surface and thus naturally model both convergence and divergence. Costa-Cabral and Burges (1994) showed that: (1) the FD8/FRho8 algorithms can still generate errors in certain types of landscapes, and (2) the DEMON algorithm tends to delineate convergent and divergent flow areas more accurately than either the D8, Rho8, or FD8/FRho8 algorithms. Holmgren (1994) and Quinn et al. (1995) recently re-examined the multiple flow direction algorithms and suggested several modifications that may eliminate some of these problems. The fraction of contributing area passed from a cell to neighbor *i* in the initial multiple flow direction algorithms was given by:

$$f_i = S^p \,/\, \Sigma S^p \qquad (7)$$

where *S* is the slope from the central node to neighbor *i* and *p* is a positive constant. Freeman (1991) found that p = 1.1 produced the most accurate results for artificial conical surfaces and TAPES-G uses that value (Gallant and Wilson 1996). Recently, Holmgren (1994) reported that much higher values of *p* ranging from 6 to 8 may be more appropriate for many natural landscapes, and Quinn et al. (1995) showed that raising the power in (7) tends to give more of a single flow direction (Figure 8). The optimum solution will vary with the type of landscape and the grid size that is chosen.

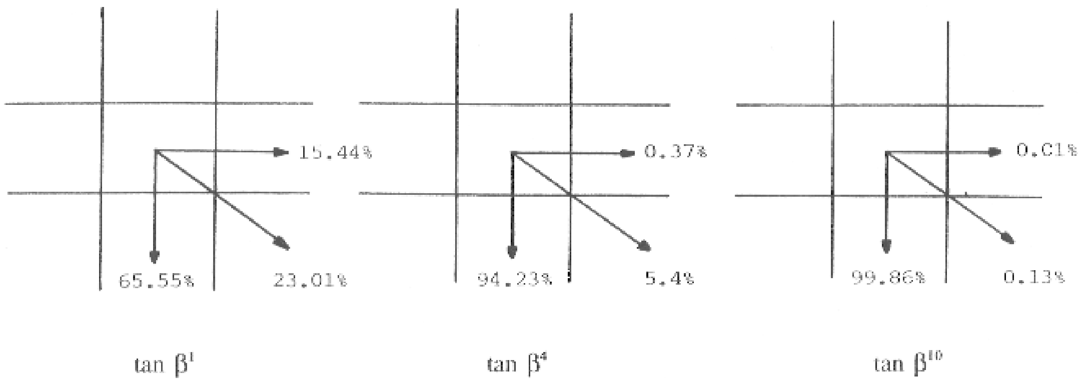


Figure 8. Effect of the flow apportioning routine when raising tanß to the power h (from Quinn et al. 1995, 170)

The choice of grid size and flow apportioning method will affect the runoff dynamics in TOPMODEL (Wolock and McCabe 1994, Quinn et al. 1995). The hydrograph predictions may not be affected because the changes in the magnitude and distribution of the topographic index will be offset by changes to the soil hydraulic conductivity and lateral transmissivity terms when the model is re-optimized. However, these changes will affect the pattern of variable source areas. These effects are complicated because channels represent subgrid features in many upland catchments and the methods used to handle the cells containing channels will have important consequences for the delineation of variable source areas in and around channels and channel heads (Quinn et al. 1995).

Morris and Heerdegen (1989) used the upslope contributing area as a threshold to initiate channels. Gallant and Wilson (1996) refer to this threshold as the maximum cross grading area in TAPES-G and they use it to switch from a multiple flow algorithm (in upland areas) to a single flow algorithm (in channel cells) for routing water to the catchment outlet. Quinn et al. (1995) have taken this idea further and suggested a new method for calculating the topographic index in grid cells containing a channel. They also compared the topographic index distributions for a series of successively smaller values of this channel initiation threshold (CIT) and found that unrealistic variable source area patterns were generated if the CIT was set incorrectly. They asked whether there is an optimum threshold for channel initiation and proposed using the CIT value which causes a rapid change in the magnitude and position of the peak value of the topographic index distribution function as this optimum value. Quinn et al. (1995) acknowledged that this CIT value may not represent the true position of the channels as measured in the field (cf. Dietrich et al. 1993, Prosser and Dietrich 1995), but argued that such a value could be defined for a particular grid resolution and that it was consistent with the aims of TOPMODEL and the assumptions of the analyses.

Quinn et al. (1995) also introduced a combined CIT and $(\tan ß)^h$ index which assumed that: (1) the permanent channel can be identified, (2) the power factor used for flow apportioning starts with the full multiple flow direction algorithm near the catchment divide (i.e. *h* is equal to unity); and (3) this term changes in concert with the upslope contributing area to generate progressively straighter flows nearer the permanent channel. They proposed the following equation to model this downslope feedback effect:

$$d_i = cld \tan\beta^f \qquad (8)$$

h

where $f = [(A/thresh) + 1]$, A = the current upslope area, *thresh* = the current CIT value chosen by the user, h = the adjustable power term, and *cld* = contour length. At low *A/thresh* the resulting power will be close to one and the multiple flow direction algorithm will be invoked. As the cells approach the specified CIT value a straighter flow direction will occur. The degree of straightening will depend on the value chosen for h and higher values of h will generate a faster change from multiple to single flow direction accumulation and tend to produce a stronger contrast in topographic index maps. Quinn et al. (1995) suggested choosing values for *thresh* and h that matched the pattern of the topographic index with field observations.

NEW POTENTIAL FOR NEW MODELS?

The descriptions of GIS-based modeling applications in the previous section offer little evidence that GIS software and databases have led to the development of new and/or improved land surface/subsurface models during the past decade. Most of the applications have simply used the GIS as a way to organize model inputs and display model predictions. The continued development of new terrain analysis methods for routing water flow across the landscape in conjunction with TOPMODEL represents a notable exception to this trend because the increasing availability of color displays afforded by GIS has accelerated the development and evaluation of these methods and led to improvements in TOPMODEL as well as the techniques themselves (Quinn et al. 1995). However, the relatively slow implementation of these new terrain analysis tools in commercial GIS packages (e.g. ARC/INFO still uses the classical D8 algorithm that represents the least preferred of the currently available algorithms) limits their availability and the opportunity to investigate their relevance to other field sites (and models).

Many of the GIS-based applications do show how difficult it is to generate reliable, location-specific estimates for key input variables and that the outputs of these models are very sensitive to small changes in the values of these input variables (e.g. De Roo et al. 1989, Brown et al. 1993, Wilson et al. 1996). This result suggests that additional work is needed to characterize the spatial variability of specific processes and properties occurring in landscapes. Two examples (one affecting the *R* factor and the other affecting the *LS* factor in the RUSLE) are reviewed here to illustrate recent developments and how these types of innovations might contribute to the development of improved land surface/subsurface models in the future.

The first example follows the proposal of Moore and Hutchinson (1991) and uses an index approach to characterize the spatial variability of the major hydrological and terrain factors affecting erosion. Moore and Wilson (1992, 1994) developed a simple dimensionless sediment transport capacity index (T_c) to predict the spatial pattern of soil loss potential that can be written as:

$$T_c = [\Sigma (\mu_i a_i / b_j) / 22.13]^m (\sin\beta_j / 0.0896)^n \qquad (9)$$

where μ_i is a weighting coefficient ($0 \leq \mu_i \leq 1$) that is dependent on the runoff generation mechanism and soil properties (i.e. infiltration rates), a_i is the area of the ith cell, b_j is the width of each cell, $ß_j$ is the slope in degrees, m and n are constants (0.6 and 1.3, respectively), and the i subscript refers to the set of i element indices which are hydrologically connected to cell j.

This equation was derived from the transport capacity limiting sediment flux in the Hairsine-Rose (Hairsine and Rose 1991, 1992a, 1992b), WEPP (Laflen et al. 1991a, 1991b), and catchment evolution (Willgoose et al. 1991) erosion theories (Moore et al. 1992). The index is equivalent to the length-slope factor in the Revised Universal Soil Loss Equation (Renard et al. 1991) for a two-dimensional hillslope, but it is simpler to use and conceptually easier to understand (Moore and Burch 1986a, 1986b). The index also distinguishes net erosion/deposition areas and accounts for different runoff producing mechanisms and soil properties using a spatially variable weighting function and it could be easily implemented within a raster-based GIS (Wilson and Gallant 1996).

The second example provides an alternative to the EI_{30} index used for the R factor in the USLE and RUSLE models. The product of storm rainfall kinetic energy (E) and the maximum rainfall intensity measured over 30 minutes (I_{30}) has been used as the R factor in the USLE throughout its history. However, this EI_{30} index does not consider variations in runoff rate that may result from differences in infiltration rates (Figure 9) and cause variations in erosion rates similar to those illustrated in Figure 10. Kinnell et al. (1995) examined erosion data from non-vegetated plots at Holly Springs, Mississippi and demonstrated that their $I_X E_A$ index is superior to the EI_{30} index because it accounts for the processes of detachment and transport better than the EI_{30} index. This $I_X E_A$ index, first proposed by Kinnell (1983), is based on the product of the excess rainfall rate (I_X) and the rate of expenditure of rain kinetic energy (E_A). Rainfall kinetic energies are computed from:

$$E_A = 1099I\,[1 - 0.72\exp(-1.27I)] \qquad (10)$$

where E_A has units of ft-ton per acre per hour and I is rainfall intensity in inches per hour. This equation follows directly from the intensity-unit kinetic energy relationship used by Renard et al. (1993) in RUSLE (Kinnell et al. 1995). The excess rainfall term (I_X), which serves as a surrogate for the runoff rate (Q), is the difference between the rainfall intensity and infiltration rate (I_S). Spatially-variable estimates of I_X (and therefore the $I_X E_A$ index) could be obtained from a soils database (that included the infiltration rate as an attribute) and monthly values of I and/or I_S could be used in place of the EI_{30} index to predict the effects of different crops and cultivation practices on soil erosion (similar to the current model).

The $I_X E_A$ index is similar to the T_c index of Moore and Wilson (1992, 1994) in that it provides an opportunity to consider the effects of hydrology more directly within the USLE/RUSLE modeling environment than is currently possible (Kinnell et al. 1995). They are related to the extent that the $I_X E_A$ index can be used to help assign weights in (*9*) for landscapes dominated by Hortonian (infiltration-excess) overland flow. However, additional research is needed to evaluate the effects of using these alternative *R* and/or *LS* factors on *K* (which incorporates the effects of profile permeability) and the other RUSLE factors.

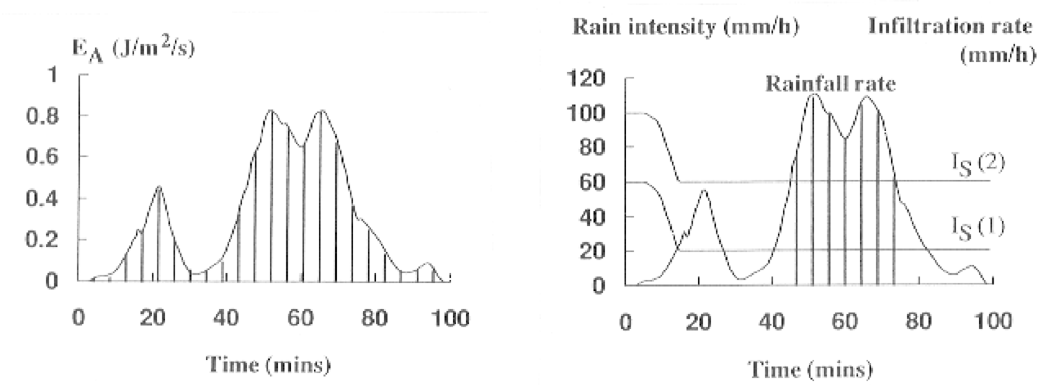


Figure 9. Schematic representation of EI_{30} index in relation to temporal variations in rainfall kinetic energy flux (E_A) and rainfall intensity. The EI_{30} for the event is given by twice the product of the shaded areas in Parts A and B, and examples of temporal variations in soil infiltration rate (I_S(1), I_S(2)) that might apply under these circumstances are also shown in Part B (from Kinnell et al. 1995, 1450).

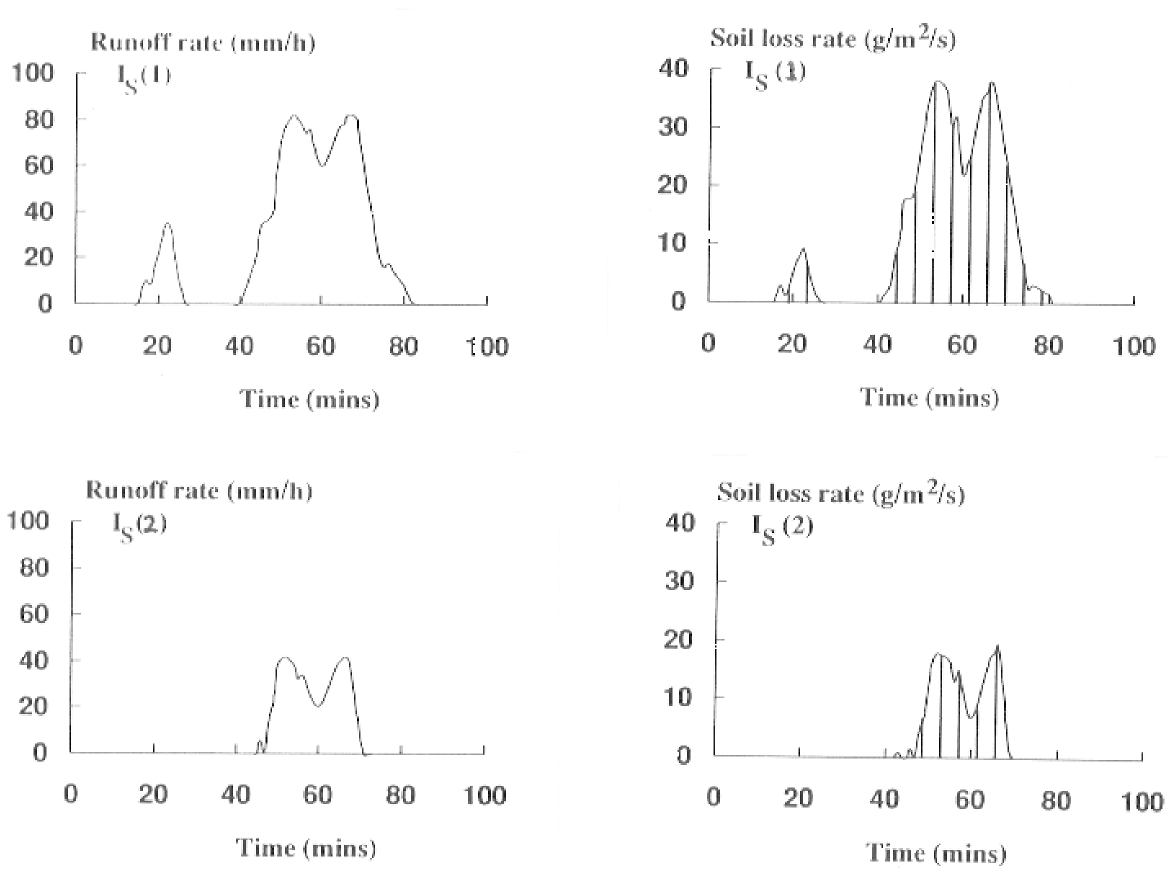


Figure 10. Possible variations in runoff and soil erosion resulting from the rainfall and infiltration conditions given in Figure 9 (from Kinnell et al. 1995, 1450)

The $I_X E_A$ index is similar to the *Tc* index of Moore and Wilson (1992, 1994) in that it provides an opportunity to consider the effects of hydrology more directly within the USLE/RUSLE modeling environment than is currently possible (Kinnell et al. 1995). They are related to the extent that the $I_X E_A$ index can be used to help assign weights in (*9*) for landscapes dominated by Hortonian (infiltration-excess) overland flow. However, additional research is needed to evaluate the effects of using these alternative *R* and/or *LS* factors on *K* (which incorporates the effects of profile permeability) and the other RUSLE factors.

The validation of GIS-based models may require additional data at the grid scale because grid-scale measurements may not exist for some model outputs and they may not be adequately described by point measurements in others. Repetto and Wilson (1996) recently computed the sediment transport index for the Palouse study site of Busacca et al. (1993) and found that the index could not predict the spatial pattern of net erosion and net deposition (see Figure 1 for diagrams showing patterns inferred from Cs-137 measurements). This result might be attributed to: (1) the failure of the model (and therefore the importance of factors other than topography in controlling the spatial patterns of erosion and deposition) (Quine and Walling 1993); (2) the failure of the 30 m DEM that was used as source data to capture the variability in the topographic form of the study area (see Zhang and Montgomery (1994) and Quinn et al. (1995) for similar claims), and/or (3) the failure of the Cs-137 point measurements to accurately represent the spatial patterns of soil erosion and deposition occurring at this level of spatial aggregation.

Overall, the GIS-based modeling experiences of the past ten years reiterate the need to develop new methods for collecting and characterizing the spatial variability of key processes and properties in landscapes, and the importance of modeling error and uncertainty in spatial databases and their effects on model predictions. Geographic information systems and related technologies (GPS receivers, remote sensing platforms, geostatistical techniques, etc.) can help with the collection and interpretation of these data and by doing so expedite the development of new and improved spatial models of key land surface/subsurface processes in future years.

REFERENCES CITED

Addiscott, T.M. (1977) A simple computer model for leaching in structured soils. *Journal of Soil Science* 28: 544-563.

Beasley, D.B., and Huggins, L.F. (1982) *ANSWERS (Areal Nonpoint Source Watershed Environmental Response Simulation) User's Manual* . Chicago: U.S. Environmental Protection Agency Report No. 905/9-82-001.

Beasley, D.B., Huggins, L.F., and Monke, E.J. (1982) Modeling sediment yields for agricultural watersheds. *Journal of Soil and Water Conservation* 37(2): 113-117.

Beven, K.J. (1989) Changing ideas in hydrology - the case of physically-based models. *Journal of Hydrology* 105: 157-172.

Beven, K.J., and Kirkby, M.J. (1979) A physically-based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin* 24: 43-69.

Beven, K.J., Kirkby, M.J., Schofield, N., and Tagg, A.F. (1984) Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments. *Journal of Hydrology* 69: 119- 143.

Binley, A., Egly, J, and Beven, K.J. (1989) A physically-based model of heterogeneous hillslopes, 1. Runoff production, 2. Effective hydraulic conductivities. *Water Resources Research* 25: 1219- 1233.

Bliss, N.B., and Reybold, W.U. (1989) Small-scale digital soil maps for interpreting natural resources. *Journal of Soil Water Conservation* 44(1): 30-34.

Brown, D.G., Bian, L., and Walsh, S.J. (1993) Response of a distributed watershed model to variations in input data aggregation levels. *Computers and Geosciences* 19(4): 499-509.

Busacca, A.J., Cook, C.A., and Mulla, D.J. (1993) Comparing landscape-scale estimation of soil erosion in the Palouse using Cs-137 and RUSLE. *Journal of Soil and Water Conservation* 48(4): 361- 367.

Costa-Cabral, M., and Burges, S.J. (1994) Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. *Water Resources Research* 30(6): 1681-1692.

Custer, S.G., Farnes, P.E., Wilson, J.P., and Snyder, R.D. (1996) A comparison of hand and spline-drawn precipitation maps for mountainous Montana. *Water Resources Bulletin* 32(2): in press.

Daly, C., Neilson, R.P., and Phillips, D.L. (1994) A statistical-topographic approach to modeling the distribution of precipitation in mountainous terrain. *Journal of Applied Meteorology* 33(1): 140-158.

De Roo, A.P.J., Hazelhoff, L., and Burrough, P.A. (1989) Soil erosion modelling using ANSWERS and geographical information systems. *Earth Surface Processes and Landforms* 14: 517-532.

Dietrich, W.E., Wilson, C.J., Montgomery, D.R., and McKean, J. (1993) Analysis of erosion thresholds, channel networks, and landscape morphology using a digital terrain model. *Journal of Geology* 101(2): 259-278.

Engel, B.A., Srinivasan, R., and Rewerts, C. (1993) A spatial decision support system for modeling and managing agricultural non-point source pollution. pp. 231-237 in Goodchild, M.F., Parks, B.O. and Steyaert, L.T., eds., *Environmental Modeling with GIS*. New York: Oxford University Press.

Fairfield, J., and Leymarie, P. (1991) Drainage networks from grid digital elevation models. *Water Resources Research* 27(5): 709-717.

Foster, G.R., and Wischmeier, W.H. (1974)Evaluating irregular slopes for soil loss prediction. *Transactions of the American Society of Agricultural Engineers* 17(2): 305-309.

Foster, G.R., Lane, L.J., Nowlin, J.D., Laflen, J.M., and Young, R.A. (1981) Estimating erosion and sediment yield on field-sized areas. *Transactions of the American Society of Agricultural Engineers* 24(5): 1253-1262.

Foussereau, X., Hornsby, A.G., and Brown, R.B. (1993) Accounting for variability within map units when linking a pesticide fate model to soil survey. *Geoderma* 60:257-276.

Fraser, R.H., Warren, M.V., and Barten, P.K. (1995) Comparative evaluation of land cover data sources for erosion prediction. *Water Resources Bulletin* 31(6): 991-1000.

Freeman, G.T. (1991) Calculating catchment area with divergent flow based on a regular grid. *Computers and Geosciences* 17(3): 413-422.

Frere, M.H., Ross, J.D., and Lane, L.J. (1980) The nutrient submodel. pp. 65-87 in Knisel, W.G., ed., *CREAMS: A Field Scale Model for Chemicals, Runoff, and Erosion from Agricultural Management Systems* . Washington, D.C.: U.S. Department of Agriculture, Agricultural Research Service Conservation Research Report No. 26.

Gallant, J.C., and Wilson, J.P. (1996) TAPES-G: A grid-based terrain analysis program for the environmental sciences. *Computers and Geosciences* 22(4): in press.

Garbrecht, J., and Martz, L. (1994) Grid size dependency of parameters extracted from digital elevation models. *Computers and Geosciences* 20(1): 85-87.

Glanz, J. (1994) New soil erosion model erodes farmers' patience. *Science* 264: 1661-1662.

Goodchild, M.F., Parks, B.O., and Steyaert, L.T., eds. (1993) *Environmental Modeling with GIS* . New York: Oxford University Press.

Goodchild, M.F., Steyaert, L.T., Parks, B.O., Crane, M.P., Johnston, C.A., Maidment, D.R., and Glendinning, S., eds. (1996) *GIS and Environmental Modeling: Progress and Research Issues*. Fort Collins: GIS World, Inc.

Goodrich, D.C., and Woolhiser, D.A. (1991) Catchment hydrology. *Reviews of Geophysics, Supplement: U.S. National Report to the International Union of Geodesy and Geophysics* , 1987-1990, pp. 202-209.

Grayson, R.B., Moore, I.D., and McMahon, T.A. (1992) Physically-based hydrologic modeling: II. Is the concept realistic? *Water Resources Research* 26(10): 2659-2666.

Griffin, M.L., Beasley, D.B., Fletcher, J.J., and Foster, G.R. (1988) Estimating soil loss on topographically nonuniform field and farm units. *Journal of Soil and Water Conservation* 43, 326-331.

Hairsine, P.B., and Rose, C.W. (1991) Rainfall detachment and deposition: Sediment transport in the absence of flow-driven processes. *Soil Science Society of America Journal* 55(2): 320-324.

Hairsine, P.B., and Rose, C.W. (1992a) Modelling water erosion due to overland flow using physical principles: I. Sheet flow. *Water Resources Research* 28(1): 237-243.

Hairsine, P.B., and Rose, C.W. (1992b)Modelling water erosion due to overland flow using physical principles: II. Rill flow. *Water Resources Research* 28(1): 245-250.

Hession, W.C., and Shanholtz, V.O. (1988) A geographic information system for targeting nonpoint- source agricultural pollution. *Journal of Soil and Water Conservation* 43(3): 264-266.

Holmgren, P. (1994) Multiple flow direction algorithms for runoff modelling in grid-based elevation models: An empirical evaluation. *Hydrologic Processes* 8: 327-334.

Hornberger, G.M., Beven, K.J., Cosby, B.J., and Sappington, D.E. (1985) Shenandoah watershed study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment. *Water Resources Research* 21: 1841-1850.

Hutchinson, M.F. (1989) A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* 106: 211-232.

Hutchinson, M.F. (1995) Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information Systems* 9(4): 385-403.

Hutson, J.L. (1993) Applying one-dimensional deterministic chemical fate models on a regional scale. *Geoderma* 60: 201-212.

Hutson, J.L., and Wagenet, R.J. (1993) A pragmatic field-scale approach for modeling pesticides. *Journal of Environmental Quality* 22: 494-499.

Inskeep, W.P., Wraith, J.M., Wilson, J.P., Snyder, R.D., and Macur, R.E. (1996) Input parameter and model resolution effects on solute transport predictions. *Journal of Environmental Quality* 25(3): in press.

Iorgulescu, I., and Jordan, J.-P. (1994) Validation of TOPMODEL on a small Swiss catchment. *Journal of Hydrology* 159: 255-273.

James, D.E., and Hewitt, M.J. (1992) To save a river: Building a resource decision support system for the Blackfoot River drainage. *GeoInfo Systems* 2(10): 36-49.

Joao, E.M., and Walsh, S.J. (1992) GIS implications for hydrologic modeling: Simulation of nonpoint pollution generated as a consequence of watershed development scenarios. *Computers, Environment, and Urban Systems* 16(1): 43-63.

Kinnell, P.I.A. (1983) The effect of kinetic energy of excess rainfall on soil loss from non-vegetated plots. *Australian Journal of Soil Research* 21(4): 445-453.

Kinnell, P.I.A., McGregor, K.C., and Rosewell, C.J. (1995) The IxEa index as an alternative to the EI30 erosivity index. *Transactions of the American Society of Agricultural Engineers* 37(5): 1449-1156.

Laflen, J.M., Elliot, W.J., Simanton, J.R., Holzhey, C.S., and Kohl, K.D. (1991a) WEPP: Soil erodibility experiments for rangeland and cropland soils. *Journal of Soil and Water Conservation* 46(1): 39-44.

Laflen, J.M., Lane, L.J., and Foster, G.R. (1991b) WEPP: A new generation of erosion prediction technology. *Journal of Soil and Water Conservation* 46(1): 34-38.

Lane, L.J. (1982) Development of a procedure to estimate runoff and sediment transport in ephemeral streams. pp. 275-282 in Walling, D.E., ed., *Recent Developments in the Explanation and Prediction of Erosion and Sediment Yield* . Wallingford: International Association of Hydrological Sciences Publication No. 137.

Moore, I.D. (1996) Hydrologic modeling and GIS. pp.143-148 in Goodchild, M.F., Steyaert, L.T., Parks, B.O., Crane, M.P., Johnston, C.A., Maidment, D.R., and Glendinning, S., eds., *GIS and Environmental Modeling: Progress and Research Issues*. Fort Collins: GIS World, Inc.

Moore, I.D., and Burch, G.J. (1986a) Physical basis of the length-slope factor in the Universal Soil Loss Equation. *Soil Science Society of America Journal* 50(5): 1294-1298.

Moore, I.D., and Burch, G.J. (1986b) Modelling erosion and deposition: Topographic effects. *Transactions of the American Society of Agricultural Engineers* 29(6): 1624-1630, 1640.

Moore, I.D., and Gallant, J.C. (1991) Overview of hydrologic and water quality modeling. pp. 1-8 in Moore, I.D., ed., *Modeling the Fate of Chemicals in the Environment* . Canberra: Centre for Resource and Environmental Studies, Australian National University.

Moore, I.D., and Wilson, J.P. (1992) Length-slope factors for the Revised Universal Soil Loss Equation: Simplified method of estimation. *Journal of Soil and Water Conservation* 47(5): 423-428.

Moore, I.D., and Wilson, J.P. (1994) Reply to "Comment on Length-slope factors for the Revised Universal Loss Equation: Simplified method of estimation" by George R. Foster. *Journal of Soil and Water Conservation* 49(2): 174-180.

Moore, I.D., Lewis, A., and Gallant, J.C. (1993a) Terrain attributes: Estimation methods and scale effects. pp. 189-

214 in Jakeman, A.J., Beck, M.B., and McAleer, M., eds, *Modelling Change in Environmental Systems*. New York: John Wiley and Sons.

Moore, I.D., Turner, A.K., Wilson, J.P., Jenson, S.K., and Band, L.E. (1993b) GIS and land surface-subsurface modeling. pp. 196-230 in Goodchild, M.F., Parks, B.O. and Steyaert, L.T., eds, *Environmental Modeling with GIS*. New York: Oxford University Press.

Morris, D.M., and Heerdegen, R.G. (1988) Automatically derived catchment boundaries and channel networks and their hydrological applications. *Geomorphology* 1(2): 131-141.

Nicholls, P.H., Walker, A., and Baker, R.J. (1982) Measurement and simulation of the movement and degradation of atrazine and metribuzin in a fallow soil. *Pesticide Science* 13: 484-494.

Nielsen, G.A., Caprio, J.M., McDaniel, P.A., Snyder, R.D., and Montagne, C. (1990) MAPS: A GIS for land resource management in Montana. *Journal of Soil and Water Conservation* 45(4): 450-453.

Nofziger, D.L., and Hornsby, A.G. (1986) A microcomputer-based management tool for chemical movement in soil. *Applied Agricultural Research* 1(1): 50-56.

Nofziger, D.L., and Hornsby, A.G. (1987) *Chemical Movement through Layered Soils Model Users Manual* . Gainesville: Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida

Obled, Ch., Wendling, J., and Beven, K.J. (1994) The sensitivity of hydrological models to spatial rainfall patterns: An evaluation using observed data. *Journal of Hydrology* 159: 305-333.

O'Callaghan, J.F., and Mark, D.M. (1984) The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics and Image Processing* 28: 323-344.

Panuska, J.C., Moore, I.D., and Kramer, L.A. (1991) Terrain analysis: Integration into the agricultural nonpoint source (AGNPS) pollution model. *Journal of Soil and Water Conservation* 46(1): 59- 64.

Pennell, K.D., Hornsby, A.G., Jessop, R.E., and Rao, P.S.C. (1990) Evaluation of five simulation models for predicting aldicarb and bromide behavior under field conditions. *Water Resources Research* 26: 2679-2693.

Petach, M.C., Wagenet, R.J., and DeGloria, S.D. (1991) Regional water flow and pesticide leaching using simulations with spatially distributed data. *Geoderma* 48: 245-269.

Prosser, I.P., and Dietrich, W.E. (1995) Field experiments on erosion by overland flow and their implication for a digital terrain model of channel initiation. *Water Resources Research* 31(11): 2867-2876.

Quine, T.A., and Walling, D.E. (1993) Use of caesium-137 measurements to investigate relationships between erosion rates and topography. pp. 31-48 in Thomas, D.S.G., and Allison, R.J., eds., *Landscape Sensitivity*. London: John Wiley and Sons.

Quinn, P.F., and Beven, K.J. (1993) Spatial and temporal predictions of soil moisture dynamics, runoff, variable source areas and evapotranspiration for Plynlimon, Mid-Wales. *Hydrological Processes* 7: 425-448.

Quinn, P.F., Beven, K.J., Chevallier, P., and Planchon, O. (1991) The prediction of hillslope paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes* 5(1): 59-79.

Quinn, P.F., Beven, K.J., and Lamb, R. (1995) The ln(a/tanß) index: How to calculate it and how to use it within the TOPMODEL framework. *Hydrological Processes* 9: 161-182.

Rawls, W.J., and Brakensiek, D.L. (1982) Estimating soil water retention from soil properties. *Journal of the Irrigation Division, American Society for Civil Engineers* 108: 166-171.

Renard, K.G., Foster, G.R., Weesies, G.A., and Porter, J.P. (1991) RUSLE: Revised Universal Soil Loss Equation. *Journal of Soil and Water Conservation* 46(1): 30-33.

Renard, K.G., Foster, G.R., Weesies, G.A., McCool, D.K., and Yoder, D.C. (1993) Predicting Soil *Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation* . Washington, D.C.: U.S. Department of Agriculture, Agriculture Handbook No. 703.

Repetto, S., and Wilson, J.P. (1996) Identification of areas experiencing net erosion and deposition. In Wilson, J.P., and Gallant, J.C., eds, *Terrain Analysis Methods for the Environmental Sciences*. Cambridge: GeoInformation International (forthcoming).

Reybold, W.U., and TeSelle, G.W. (1989) Soil geographic data bases. *Journal of Soil and Water Conservation* 44(1): 28-29.

Richardson, C.W., and Wright, D.A. (1984)*WGEN: A model for generating daily weather variables* . Washington, D.C.: United States Department of Agriculture Report No. ARS-8.

Ritchie, J.C., and McHenry, J.R. (1990) Application of radioactive fallout Cesium-137 for measuring soil erosion and sediment accumulation rates and patterns: A review. *Journal of Environmental Quality* 19: 215-233.

Robson, A.J., Beven, K.J., and Neal, C. (1992) Towards identifying sources of subsurface flow: A comparison of components identified by a physically based runoff model and those determined by mixing techniques. *Hydrological Processes* 6: 199-214.

Robson, A.J., Whitehead, P.G., and Johnson, R.C. (1993)An application of a physically based semi-distributed model to the Balquhidder catchments. *Journal of Hydrology* 145: 357-370.

Romanowicz, R., Beven, K.J., and Moore, R. (1993) TOPMODEL as an application module within WIS. pp. 211-223 in Kovar, K. and Nachpaecel, H.P., eds., *Applications of Geographic Information Systems in Hydrology and Water Resources*. Wallingford: International Association of Hydrological Sciences Publication No. 211.

Running,S.W. and P.E. Thornton (1996) Generating daily surfaces of temperature and precipitation over mountainous terrain. pp. 93-98 in Goodchild, M.F., Steyart,L.T., Parks,B.O., Crane,M.P., Johnston,C.A., Maidment,D.R., and Glendinning,S. eds., *GIS and Environment Modeling: Progress and Research Issues*. Fort Collins: GIS World, Inc. (forthcoming).

Sivapalan, M., Beven, K.J., and Wood, E.F. (1987) On hydrologic similarity: 2, A scaled model of storm runoff production. *Water Resources Research* 23: 2266-2278.

Spangrud, D.J., Wilson, J.P., Nielsen, G.A., Jacobsen, J.S., and Tyler, D.A. (1995) Sensitivity of computed terrain attributes to the number and pattern of GPS-derived elevation data. pp. 285- 301 in Robert, P.C., Rust, R.H., and Larson, W.E., eds., *Site-Specific Management for Agricultural Systems*. Madison: American Society of Agronomy.

Stephens, P.R., MacMillan, J.H., Daigle, J.L., and Chilar, J. (1985) Estimating universal soil loss equation factor values with aerial photography. *Journal of Soil and Water Conservation* 40(1): 293-296.

Tomlin, C.D. (1980) *The Map Analysis Package*. New Haven: School of Forestry and Environmental Science, Yale University.

Ventura, S.J., Chrisman, N.R., Connors, K., Gurda, R.F., and Martin, R.W. (1988) A land information system for soil erosion control planning. *Journal of Soil and Water Conservation* 43(3): 230-233.

Wagenet, R.J., and Hutson, J.L. (1986) Predicting the fate of non-volatile pesticides in the unsaturated zone. *Journal of Environmental Quality* 15: 315-322.

Wagenet, R.J., and Hutson, J.L. (1989) *LEACHM: Leaching Estimation and Chemistry Model - A process based model of water and solute movement, transformations, plant uptake, and chemical reactions in the unsaturated zone*. Ithaca: Water Resources Institute, Cornell University.

Wagenet, R.J., and Hutson, J.L. (1996) Scale dependency of solute transport modeling/GIS applications. *Journal of Environmental Quality* 25(3): in press.

Wagenet, R.J., Hutson, J.L., and Biggar, J.W. (1989) Simulating the fate of a volatile pesticide in unsaturated soil:

A case study. *Journal of Environmental Quality* 18: 78-83.

Willgoose, G., Bras, R.L., and Rodriguez-Iturbe, I. (1991) A coupled channel network growth and hillslope evolution model: 1. Theory. *Water Resources Research* 27(7): 1671-1684.

Wilson, J.P. (1986) Estimating the topographic factor in the universal soil loss equation for watersheds. *Journal of Soil and Water Conservation* 41(3): 179-184.

Wilson, J.P. (1989) Soil erosion from agricultural land in the Lake Simcoe-Couchiching Basin, 1800-1981. *Canadian Journal of Soil Science* 69(2): 206-222.

Wilson, J.P. (1996) Spatial models of land use systems and soil erosion: The role of GIS. In Wegener, M., and Fotheringham, A.S. eds., *GIS and Spatial Models: New Potential for New Models?* London: Taylor and Francis (forthcoming).

Wilson, J.P., and Gallant, J.C. (1996) EROS: A grid-based program for estimating spatially-distributed erosion indices. *Computers and Geosciences* 22(4): in press.

Wilson, J.P., Inskeep, W.P., Rubright, P.R., Cooksey, D., Jacobsen, J.S., and Snyder, R.D. (1993) Coupling geographic information systems and models for weed control and groundwater protection. *Weed Technology* 7(1): 255-264.

Wilson, J.P., Inskeep, W.P., Wraith, J.M., and Snyder, R.D. (1996) GIS-based solute transport modeling applications: Scale effects of soil and climate databases. *Journal of Environmental Quality* 25(3): in press.

Wischmeier, W.H. (1976) Use and misuse of the universal soil loss equation. *Journal of the Soil and Water Conservation* 31(1): 5-9.

Wischmeier, W.H., and Smith, D.D. (1978) *Predicting Rainfall Erosion Losses: A Guide to Conservation Planning* . Washington, D.C.: U.S. Department of Agriculture, Agriculture Handbook No. 537.

Wolock, D.M. (1995) Effects of subbasin size on topographic characteristics and simulated flow paths in Sleepers River watershed, Vermont. *Water Resources Research* 31(8): 1989-1997.

Wolock, D.M., and McCabe, G.J. (1994) Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research* 31(5): 1315- 1324.

Wolock, D.M., Hornberger, G.M., and Musgrove, T.M. (1990) Topographic effects on flow path length and surface water chemistry of the Llyn Brianne catchments in Wales. *Journal of Hydrology* 115:243-259.

Young, R.A., Onstad, C.A., Bosch, D.D., and Anderson, W.P. (1987) AGNPS, *Agricultural Nonpoint Source Pollution Model: A large watershed analysis tool*. Washington, D.C.: U.S. Department of Agriculture, Agricultural Research Service Conservation Research Report No. 35.

Young, R.A., Onstad, C.A., Bosch, D.D., and Anderson, W.P. (1989) AGNPS: A nonpoint source pollution model for evaluating agricultural watersheds. *Journal of the Soil and Water* Conservation 44(2): 168-173.

Young, R.A., Otterby, M.A., and Roos, A. (1982) A technique for evaluating feedlot pollution potential. *Journal of the Soil and Water Conservation* 37(1): 21-23.

Zhang, W. and D.R. Montgomery (1994) Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research* 30(4): 1019-1028.

Return to beginning of document

Wilson,J.P., GIS-based Land Surface/Subsurface Modeling: New Potential for New Models?

John P. Wilson
Geographic Information and Analysis Center
Department of Earth Sciences

Montana State University
Bozeman Montana MT 59717-0348

COMPARISON OF ANUSPLIN, MTCLIM-3D, AND PRISM PRECIPITATION ESTIMATES

by
Sara T. Stillman and John P. Wilson, Department of Earth Sciences, Montana State University
Christopher Daly, Forestry Sciences Laboratory, Oregon State University
Michael F. Hutchinson, Centre for Resource and Environmental Studies, Australian National University; and
Peter E. Thornton, School of Forestry, University of Montana.

A comparison of the ANUSPLIN, MTCLIM-3D, and PRISM model performance is needed to assist users with appropriate model selection and elucidate potential differences. The models employ different techniques to develop gridded precipitation surfaces from published climate station (point) data and digital elevation models (DEMs). Mean monthly and annual precipitation estimates were prepared for the Bozeman, Billings, Ashton, and White Sulphur Springs 1 x 2o topographic quadrangles in southwestern Montana and the Cody quadrangle in Wyoming for the 1961-90 data period to determine whether the predicted precipitation surfaces are hydrologically reasonable over a region which contains a diverse physiography and produces a wide range of precipitation regimes. Input data included mean monthly and annual precipitation data from 258 weather stations and a 0.5 km square-grid DEM derived from the appropriate 3 arc-second USGS DEMs with ANUDEM.

The models generated statistically similar results. The mean annual precipitation predictions for the 20 (randomly selected) withheld stations were accepted as statistically similar to the observed data at the 0.05 significance level. ANUSPLIN produced slightly higher mean annual estimates (5.6% and 4.0% higher than MTCLIM-3D and PRISM, respectively), and tended to overestimate precipitation at the 20 withheld stations. This model also generated slightly higher mean absolute errors (MAE) compared to the other two models which tended to underestimate precipitation at the 20 withheld stations. The largest differences between the model predictions occur in high elevation areas (i.e., Absarokas, Tetons) where a lack of climate stations and highly variable precipitation patterns complicate the interpolation process. Similar annual results suggest model selection should be based on ease of use and efficiency.

The MTCLIM-3D mean values are higher in winter and early spring while PRISM predictions are greater in the late spring-summer months and ANUSPLIN generally has the lowest predictions overall although the differences are small. The predictions fell in the same 25 mm classes in over 80% of the DEM cells in 28 out of 36 monthly surface comparisons. The largest differences occurred in late fall and winter. The largest MAE and bias estimates were generated for all three models in these months. In addition, predictions were different from the observed data at the 20 withheld stations at the 0.05 significance level in November for ANUSPLIN and MTCLIM-3D, December for ANUSPLIN and PRISM, and February for

ANUSPLIN. Relatively low agreement between the summed monthly surface and annual surface values for all three models demonstrate the importance of incorporating snow course data.

The models require less climate knowledge compared to the hand-contouring method and provide error estimates as well as precipitation estimates that can be accessed directly by modern geographic information systems. Increased numbers of climate stations in high elevations and more precise measurements of station locations and elevations would improve model predictions in the northern Rocky Mountains.

Jennifer Kesteven and Michael Hutchinson

SPATIAL MODELLING OF CLIMATIC VARIABLES ON A CONTINENTAL SCALE

This paper describes the generation of surfaces of climatic variables for the Australian Continent. The surfaces are applied to monitoring climate change and climate variability and to the modelling of the spatial distribution of additional climatic variables. Sea level surfaces of pressure and temperature for the period January 1952 to December 1990 were produced by fitting a partial thin plate smoothing spline to data from the network of Australian meteorological stations which take 3 hourly readings.

Summary statistics and graphical representation of the surfaces are used to check for data integrity and for monitoring climate change at both continental and regional scale. The monthly surfaces show that there have been some dramatic shifts in the climate on the Australian continent during the period under investigation.

Two models which derive the dew point temperature for each month are presented, a multivariate spatial regression analysis based on monthly dry and wet bulb temperature surfaces, and a lag regression model based on the previous three months dew point temperature. A third model, which integrates space and time within a spline function, is also suggested.

INTRODUCTION

Spatial analysis and modelling of climatic variables is important for economic, environmental and social reasons, especially since climate change and climate variability have become issues of global concern. Various methods to spatially interpolate sparse point observations have been used, including partitioning into regions, moving average, kriging and thin plate smoothing splines.

This paper discusses the method used to generate estimates of the spatial distribution of climatic variables. The surfaces can be used for climate diagnostics and forecasting purposes. Regular grids obtained from the surfaces are then used for the statistical analysis, and two different models to spatially represent dew point temperature are presented. The first spatial model derives the dew point temperature surface from the dry bulb and wet bulb temperature surfaces, and the second, a spatial lag regression model derives the dew point temperature surface from surfaces for the previous three months. A model which the integrates space and time within a spline function is also posed.

SPATIAL INTERPOLATION WITH THIN PLATE SMOOTHING SPLINES

Various methods to spatially interpolate sparse observational data have been used including partitioning into regions, moving averages, kriging and thin plate smoothing splines. Of these methods, statistical interpolation techniques appear to be best suited to the task. The techniques include the methods of thin plate smoothing splines (Wahba and Wendelberger 1980, Hutchinson 1991) and kriging (Delfiner and Delhomme 1975, Cressie 1991). These techniques model spatial distribution as a function of observational data across a region without prior knowledge of the distribution or its underlying physical causes. Both methods attempt to achieve minimum error (optimum) interpolation and essentially have the same underlying computational structure and have measures of their spatial accuracy. There are however some significant practical and theoretical differences between the two methods which have been extensively covered by Hutchinson and Gessler (1994) and Hutchinson (1993).

While Kriging and thin plate spline methods can give rise to results of similar accuracy (Laslett 1994, Hutchinson and Gessler 1994), there is a practical difference between them. Kriging requires the spatial covariance function or variogram to be estimated first, and is critical to the process. Thin plate smoothing splines, on the other hand, require the estimation of a smoothing parameter that determines an optimal balance between fidelity to the data and smoothness of the fitted spline function. As this is normally done totally automatically by minimising the GCV (Generalized Cross Validation), the thin plate smoothing splines are easier to use.

Bi-variate interpolation of mean monthly climate is rarely appropriate. The dependence on longitude and latitude allows the surface to incorporate broad trends with respect to position. There are, however, significant additional forcing variables, in particular elevation effects on pressure and temperature. ANUSPLIN (a package of programs for fitting surfacesdeveloped by Hutchinson) allows for the incorporation of a linear parametric sub-model, which permits a known physical control of the variable to be incorporated as a dependent variable. This allows for more accurate interpolation than by other methods and is an advantage when dealing with data which are limited, either in terms of accuracy or spatial density. The semi-parametric models are known as partial thin plate splines and have been applied to the interpolation of monthly mean temperatures by Hutchinson (1991).

THE AUSTRALIAN CLIMATE SURFACES

Since both pressure and temperature decline approximately linearly with elevation, it is natural to interpolate these climatic variables by a partial spline with two independent position variables (longitude and latitude) and a single linear dependence on elevation. An advantage of the partial spline formulation is that the coefficient of the linear sub-model (the lapse rate) is determined automatically from the data, thus it does not have to be specified beforehand. This statistical surface can then be attached to a digital elevation model (DEM) if surface values are required.

Surface construction

To model the spatial distribution of a climatological variable, the climatic series for each site must be summarized. A monthly climate database was constructed from the 3 hourly records from the Australian Bureau of Meteorology for all stations for the period 1952 to 1990, for linearly interpolated specific UTC (universal coordinated time) times (0000 and 0600). These times were selected in order to accommodate the different time zones across the country (usually three) and to remove station inhomogeneity problems associated with a introduction of Daylight Saving Time (DST) in the early 1970s for some Australian states during the summer months. A monthly mean value for each station was then calculated and monthly sea level climate surfaces constructed for pressure, dry bulb temperature, wet bulb temperature and dew point temperature with a linear dependence on elevation via a partial spline model.

The Australian Bureau of Meteorology (BoM) three hourly records comprised some 2019 stations at the end of 1990, however, the majority of these stations collected measurements at only 0900 and 1500 Local Standard Time (LST). Stations were selected by accepting a minimum of 2000 three hourly records in each year. This reduced the number of stations to 131. While the distribution of these stations is sparse in some areas, it is a good representation of all the station records from the BoM.

The surface constructions begin by generating spline coefficients from the database for each month for both 0000 and 0600 UTC. The spline technique is an iterative procedure in climate surface construction and the results of the spline interpolation provide valuable information on the database. Climate data are often unevenly spatially distributed, measured for a limited duration, and can be of varying quality. The ANUSPLIN programs always produce a diagnostics file which includes summary statistics, a list of the 100 largest residuals and a list of the data and fitted values. Files with estimated standard errors and files containing the coefficients of the fitted surfaces and the parameters used to calculate the optimum smoothing parameter may also be written. The surface coefficients are used to calculate values of the fitted surface by other programs in ANUSPLIN.

There are several useful output statistics in the diagnostics file generated by the ANUSPLIN programs, which can be used to check for data homogeneity. The diagnostics file includes a signal (an estimate of the effective number of parameters in the fitted model) and error (an estimate of the effective degrees of freedom of the residual sum of squares of the fitted model) which together add to the number of data points. The surface fitting procedure is normally considered to have failed to find a genuine optimum value of the smoothing parameter if the signal is the maximum possible or the signal is the minimum possible. Both of these conditions are flagged by an asterisk in the output. The signal to error ratio is a good initial data check. Flagged months were often found to have missing data values.

The generalized cross validation (GCV), mean square residual (MSR), the error variance estimate (VAR) and the mean square error (MSE) are given together with their square roots (RTGCV, RTMSR, RTVAR, RTMSE) which are in the units of the data. The RTGCV can be interpreted as the root mean square validation error, which includes measurement error in the data. The root mean square error (RTMSE) is an estimate of the true error of the fitted surface after the effects of measurement error have been removed.

Detecting data inhomogeneity and data errors

The RTGCV is a good measure of the predictive power of the fitted surface, especially when comparing the values from the same climate variable over time. In combination with the residuals listing, which allows for each station to be assessed in relation as to its deviation from the calculated surface, the RTGCV provides a powerful tool for investigation of station homogeneity. The example given in figure 2 is a time series of the RTGCV for 0000 UTC MSL pressure for all stations in the data base.

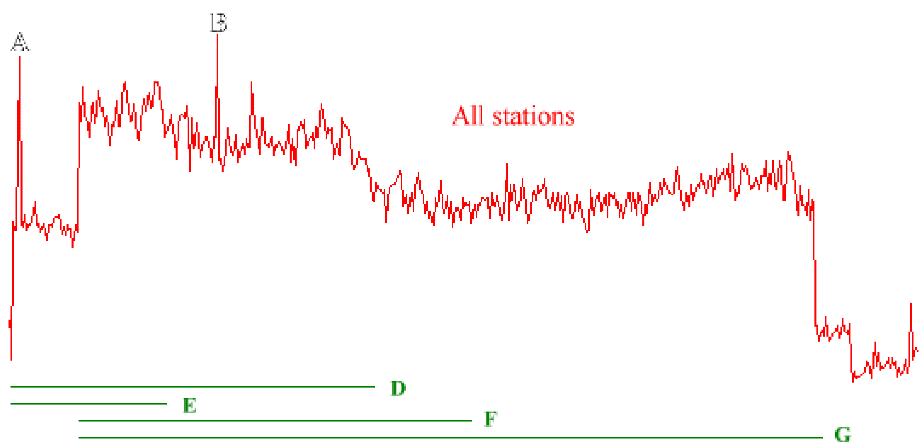


Figure 1. Time series of the RTGCV for pressure 1952-1990, all stations.

The five stations with the highest root mean square residuals in the diagnostics file for the years represented by points A (June 1952) and B (December 1960) and January 1982 are given in table 1

Table 1. Ranked root mean square residuals (hPa) for June 1952, December 1960 and January 1982 for all stations.

		Jun 52		Dec 60		Jan 82
rank	station	RTMSR	station	RTMSR	station	RTMSR
1	086038	9.58	086071	7.63	086071	8.11
2	009034	4.30	027022	5.00	040214	2.98
3	009021	2.84	009034	4.08	094029	2.25

4	017043	2.16	066062	3.91	094008	2.20
5	031011	2.12	031011	2.74	009021	2.19

Figure 2, below, shows a time series for the stations with root mean square residuals greater than three in table 1. Some errors are simply positional. For example, the station dictionary describing the longitude, latitude, and elevation can contain errors and only give present location. It is not uncommon for meteorological stations and especially regional offices to be moved. The Graphs of stations 009034, 027022, 066062 and 086071 all show discontinuities in the time series, and all are consistently listed in the ten stations with the greatest residuals for the period of their discontinuities (lines D,E,F and G in figures 1 and 2).

It is possible to use the value of the covariate estimated by the spline function to approximate the station height. For example station 086071's barometer is listed as being at 113m, the lapse rate for December 1960 was 114mb/1000m and for January 1982 it was 112mb/1000m. These figures suggest that the height was 47m in 1960 and 41m in 1982. The World Weather Records 1951-1960 lists the station's barometer height at 44 metres in 1960 and the Bureau of Meteorology's 1983 dictionary lists the station's barometer height at 38.1 metres. Considering the errors associated with fitting the function, a root mean square error of 1.2mb, these values are well within estimation limits.

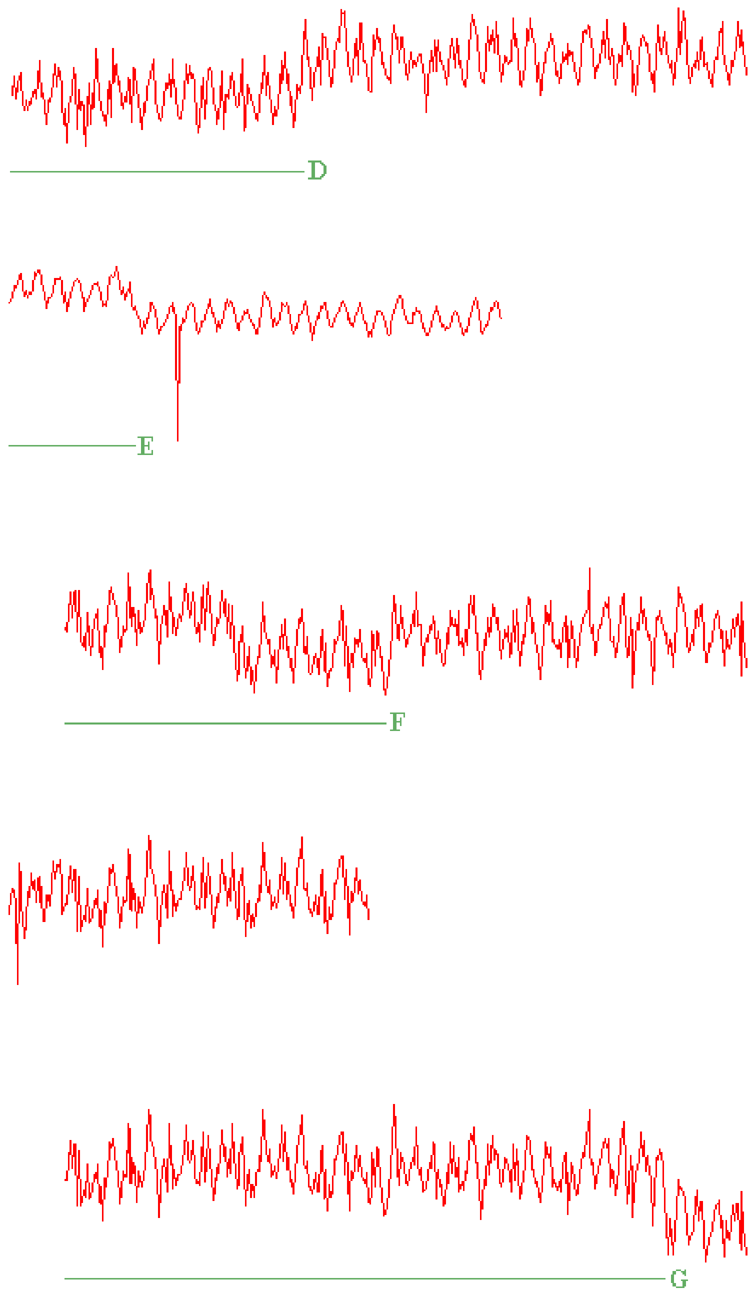


Figure 2. 0000 UTC pressure 1952-1990 for stations with high root mean square residuals

The graphs of stations 086038 and 027022 show that points A and B in figure 1 are outliers, although the latter also shows a discontinuity in the time series. These outliers arose from errors reading or recording on single days. Closer inspection of the data files reveals that on the 22nd of June 1952 station 086038 had a 0900 LST pressure recording of 5.8mb and on December 25 1960, station 027022 had a 0900 LST pressure recording of 1.2mb. When possible, errors are checked and corrected in the station database, it may however, be necessary to remove the station from the database. New coefficients are then generated. Figure 3 is a time series of the RTGCV for 0000 UTC MSL pressure with the above stations removed.

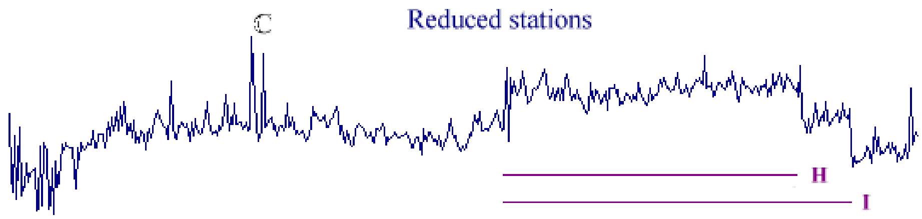


Figure 3. Time series of the RTGCV for pressure 1952-1990, reduced number of stations.

A comparison between figures 1 and 3, shows the removal of the suspect stations reduces the RTGCV from values around 2 to values around 1. Figure 3 suggests that there are still two possible outliers at point C and possibly two stations with homogeneity problems for the periods marked by lines H and I. The removal of the above stations reduces the RTMSR overall, as Table 2 shows. Some stations, such as 031011, have a much reduced error. Others, such as 040214, do not show the same improvement, and this suggests this station has data inhomogeneity problems and possibly is one of the stations suggested by H or I in figure 3.

Removing the worst stations significantly reduced the RTGCV and the root mean squared error. The signal did not change greatly, indicating robustness of the initial fit to the full data set. The iterative nature of surface generation permits careful examination of the effects of changing the station database.

Table 2: Ranked root mean square residuals (hPa) for June 1952, December 1960 and January 1982 for surfaces created with the reduced number of stations.

		Jun 52		Dec 60		Jan 82
rank	station	RTMSR	station	RTMSR	station	RTMSR
1	061078	0.82	023000	1.16	040214	2.90
2	066037	0.73	009021	1.00	009021	1.86
3	094029	0.63	029009	0.93	063231	1.58
4	031011	0.57	085072	0.73	040223	1.19
5	017043	0.54	078031	0.70	059040	1.02

This process might be repeated many times, slowly reducing overall surface error. Once the coefficients reach a level where there is very little change with the removal of additional stations, they are used to create the monthly sea level surfaces. Maps are obtained by calculating the value of the fitted function of the independent variables on a 0.1o grid of mainland Australian points at sea level. Both grid and visual representation have been created. These grids are used as the base data for the statistical analysis and modelling.

Trends in the spatial data set

The pressure and temperature surfaces have been used to produce a 0.1o grid of sea level monthly means for the Australian continent from January 1952 to December 1990 for both 0000 UTC and 0600 UTC. Monthly averages for the Australian continent were computed for each month for each variable to give a single figure representative of an Australian mean. The purpose of this was to review the behavior of Australian continental sea level pressure and temperature for the last half century based on the sea level surfaces. Figure 4 shows a 10 year running mean for pressure and temperature for 0000 UTC and 0600 UTC.

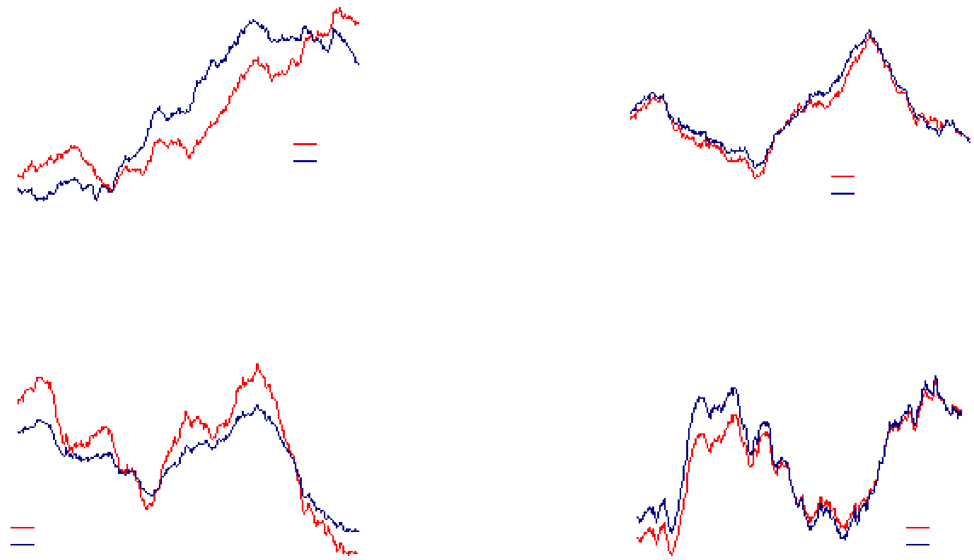


Figure 4. 10 year running means of the Australian average for temperature and pressure.

The amount of increase found in the dry bulb temperature (~0.7oC), if a true indication of the temperature trend, is alarming. An investigation of regional values of the long term dry bulb temperature showed increases in every region although some areas show no warming trend until the 1970s. This increase, although it is considerably larger, is substantiated by Jones et al (1990) who found the Southern Hemisphere area average showed little or no overall trend between about 1945 and 1970, but that since 1970 a strong warming trend has set in. They found that the overall warming trend since 1900 is about 0.5oC, of which roughly 0.3oC occurred between 1900 and 1945 and 0.2oC since 1970. It is interesting to note that the pressure, wet bulb temperature and dew point temperature, while they do not show dramatically rising trends do show a trend of a cyclic nature.

THE DEW POINT TEMPERATURE MODELS

Monthly means can be viewed as the most basic set of parameters of a statistical model of the climate and global data bases of climate data generally include only mean monthly values of climate variables. As climate accounts for a significant percentage of the variability in most ecological data, it is therefore useful to produce a climate space-time model that generates monthly climate data. Two models that use the surface data are presented. While they have been produced for the other variables in the data base, the dew point temperature models are presented here as examples. A model which integrates the space and time domains is also suggested.

Spatial linear regression model

The Spatial linear regression model is a model based on spatial regression techniques which relates the dew point temperature to dry bulb and wet bulb temperatures, As the relationship is well understood it was designed to test the spatial integrity of the spline surfaces. Relationships were derived between these variables over a calibration period (Jan 1952 to Dec 1988). The results were tested with an independent valuation data set from another period (Jan 1989 to Dec 1990), and the results for April 1989 are given.

A simple regression model relating dew point for each month at a single point over the years 1952 to 1988 to dry bulb and wet bulb temperature is given by

where Y is the dependent or response variable (dew point temperature), x1 and x2 are independent or explanatory variables (dry bulb and wet bulb temperature), alpha, beta and gamma are unknown parameters, and is the error term. This was extendextended to a spatial context by fitting equation (1) to the dry bulb, wet bulb and dew point temperatures at each one of a regular 2.5o grid of 119 points across Australia. The GENSTAT statistical package was used to determine values for alpha, beta and gamma at each of the 119 points for each month. To spatially extend the values for alpha, beta and gamma the ANUSPLIN package was then used to construct a surface for each of the parameters. Equation (1) then yields the predictive model

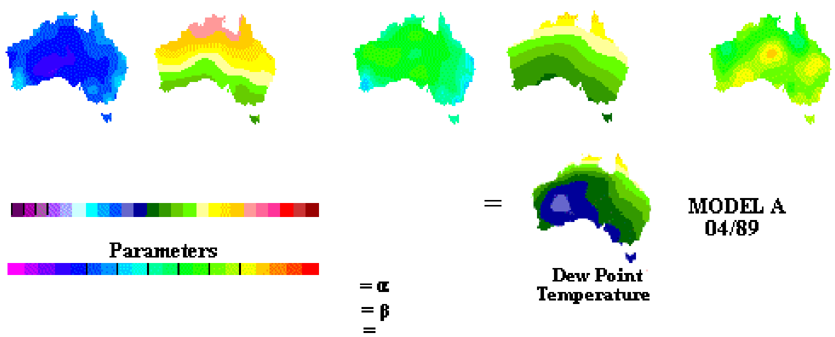


Figure 5. The dew point regession model A

While the model gives a good account of the sea level values for dew point temperature, and the percentage variance explained by the model is high (see table 3), the diagnostics from the GENSTAT statistical procedure suggested that the spatial variation in the dry bulb and wet bulb temperature

parameters are not significant. Thus the dew point model could be simplified by reducing the degrees of freedom, that is the spatially varying parameters for dry bulb and wet bulb temperature could be reduced to a non-spatially varying constant.

Table 3. Percentage variance for the dew point regression models for mid season months.

	MODEL A	MODEL B
	% variance	% variance
JAN:	99.2	99.0
APR:	97.4	97.3
JUL:	95.7	95.4
OCT:	98.3	98.0

% **variance** = 100 x (1 **-** (Residual m.s.)/(total m.s.).

The GENSTAT procedure was then rerun to estimate the values for alpha, beta and a spatially varying gamma. This gives

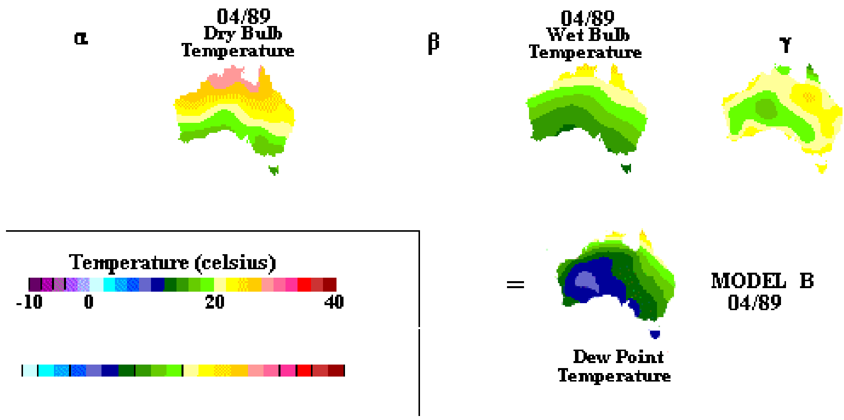


Figure 6. The dew point regession model B

This produced almost no reduction in the explanation of the percentage of the variance between models, and a comparison with the actual values shows little difference. However, over all months model A performed slightly better than model B, but at some expense in terms of its complexity. Both models however, show that the surfaces produced by the spline function display spatial integrity.

Spatial lag linear regression model

The methodology applied above was used to produce a lag regression model for monthly dew point. The regression equations were estimated for the required month as a function of the lag correlations with the dew point values for the preceeding three months and the spatially varying constant calculated using the GENSTAT statistical package. Again, April 1989 is used as an example.

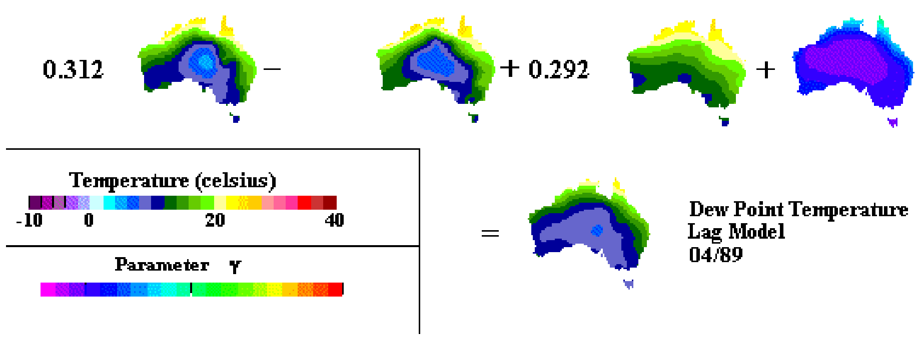


Figure 7. The dew point lag regession model

Table 4 shows that the regression model explains a high percentage of the explained variance and figure 7 shows that the model performs reasonably well for April 1989. Similar results were obtained for other months from the validation set. Both these models suggest that the spline function is well suited to the spatial extension of not only climate variables but also for the spatial extension of time series model parameters.

Table 4. Percentage variance for the dew point lag regression model for mid season months.

	% variance
JAN:	80.6
APR:	79.7
JUN:	72.7
OCT:	81.7

The space time spline model

Developments in both the previous models are guided by the requirements of the next step, the development of spline space-time model which provides the possibility of the integration of the space and time domains. The central theme underlying this model is that the relative locations of the points to which the data refer can provide some information about the spatial and temporal pattern of variation in these data. That is the data exhibit spatial and temporal correlation.

A useful starting point for a conceptual space-time model is a set of successive images of space over time. The aim here is to treat time as a independent variable in a spline function. The space-time spline model depicts processes of two-dimensional space (msl dew point temperature) that are played out along a third temporal dimension. The representation of the third (temporal) face of a changing two dimensional object is complex, due to the cyclical nature of climate. Test models which represent time as linear series have inherent problems associated with them, especially when fitting the function to a the end seasonal trend. A transformation of the time variable is required to incorporate the cyclic nature of climatic variables.

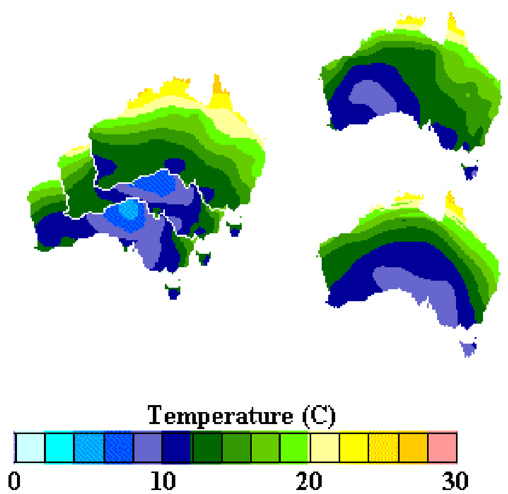


Figure 8. The space-time spline dew point model.

Transformation of the time variable to the Australian long term monthly average dew point temperature can give a marked improvement in the RMS and in the GCV. Greater improvements have been noted by using a spatially varying long term average, that is the monthly anomalies are plotted on an underlying DEM that is in effect the long term average for that month. Several tests on the model in terms of scale and order of the spline function show that better results are achieved with time as an independent variable, which suggests there is both spatial and temporal dependence. The spline function produced some unexpected results in the higher orders of the spline function, which suggest the space-time spline model is not well defined in the time domain. The approach, however, of describing time in terms of local long term mean distributions appears to hold promise of some useful outputs for monthly forecasting from limited data.

Conclusion

Central to the development of space-time models is the accurate spatial interpolation of model parameters from standard meteorological networks. By incorporating a parametric sub-model for a linear dependence on elevation, partial thin plate smoothing splines offer an objective flexible tool and an elegant and simple solution. The iterative process of fitting the spline surfaces offers an excellent tool to evaluate the integrity of the station database. In the 0000 UTC pressure surfaces there is a significant reduction in surface error after removal of the 6 stations. In this case there was no corresponding reduction in surface complexity or roughness, showing the robustness of the spline function.

Monthly surfaces have been created using the spline techniques for Australia. This database was then used as a base for the mean sea level dew point temperature models. This paper has been concerned with the development of space-time models of dew point temperature, variables such as pressure and temperature appear to have similar simple space-time structures, and we would suggest that the models would work equally well for these variables. The approach of describing time in terms of local long term mean distributions appears to hold promise, provided significant complexities in the temporal domain can be accommodated. The space-time models and the techniques used allow for the rapid revision of the surfaces as new data are entered into the station database and the models could be perturbed according to broad scale climate change scenarios produced by general circulation models.

References

Cressie N.A.C. 1991. *Statistics for Spatial Data.* New York, John Wiley and Sons.

Delfiner P. and Delhomme J.P. 1975. Optimum interpolation by kriging. In: J.D.Davis and M.J.McCullagh (eds), *Display and Analysis of Spatial Data.* New York, John Wiley and Sons, pp 96-114.

Hutchinson M.F. 1991. Continent-wide data assimilation using thin plate smoothing splines. In: J.D.Jasper (ed), *Data Assimilation Systems.* BMRC Research Report No.27, Melbourne: Bureau of Meteorology, pp 104-113.

Hutchinson M.F. 1993. On thin plate splines and kriging. In: M.E.Tarter and M.D.Lock (eds), *Computing Science and Statistics* Vol.25, Interface Foundation of North America, University of California, Berkeley, 55-62.

Hutchinson M.F. 1994. Interpolating rainfall means - getting the temporal statistics correct. *Proc. Second Inter. Conference/Workshop on GIS and Environmental Modelling.* Breckenridge, Colorado, Sept 1993.

Hutchinson M.F. and Gessler P.E. 1994. Splines - more than just a smooth interpolator. *Geoderma* 62: 45-67.

Jones P.D., Raper S.C.B., Cherry C.M., Wigley T.M.L., Santer B., Kelly P.M., Bradley R.S. and Diaz H.F. 1991. An Updated Global Grid Point Surface Air Temperature Anomaly Data Set, Environmental Sciences Division, Publication No. 3520, Oak Ridge National Laboratory.

Laslett G.M. 1994. Kriging and splines: an empirical comparison of their predictive performance in some applications. *Journal of the American Statistical Association* 89:391-409.

Wahba G. and Wendelberger J. 1980. Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review* 108:1122-1143.

Jennifer Kesteven and Michael Hutchinson,

Centre for Resource and Environmental Studies

Institute of Advanced Studies

Australian National University

Canberra ACT 0200

Australia

ph: (06)2490656

fax: (06)2490757

email jlk@cres.anu.edu.au

A Comparison of Spatial Interpolation Techniques in Temperature Estimation

- First Author: Dr. Fred C. Collins Jr.
- Affiliation: *IBM Government Systems - Resource Management and Distributed Systems*
- Address: *6300 Diagonal Highway, 003F, Boulder Colorado 80301*
- Telephone: (303) 924-4944
- FAX: (303) 924-0519
- Topic Area: GIS and Spatial Data Analysis
- Authors: Fred C. Collins, Ph.D (IBM) and Paul V. Bolstad, Ph.D.(Univ. of Minnesota)

1.0 Introduction:

The increased awareness of government and industry to the potential benefits of geographic information systems (GIS) has been driven by an increase in the availability of digital spatial data and increased hardware and software capability. Along with thi s increased awareness, has been increased concern over the accuracy and precision of spatial data. GIS error begins with data collection and continues through data input, storage, manipulation, output, and interpretation of the results. Understanding th e source, nature, and extent of errors in GIS is the first step in a strategy for reducing error in GIS. This research is concerned with the spatial interpolation of meterological data as a preliminary step prior to use in landscape, regional, and global models or as layers in a GIS.

Spatially distributed estimates of meteorological data are becoming increasingly important as inputs to spatially explicit landscape, regional, and global models. Estimates of meteorological values such as temperature, precipitation, and evapotranspirati on rate are required for a number of landscape scale models, including those of regeneration, growth, and mortality of forest ecosystems. To calculate daily microclimate conditions in mountainous terrain, the model MT-CLIM requires minimum and maximum da ily temperature data as inputs (Running and Nemani, 1987). To compute forest evapotranspiration, landscape scale ecological models such as FOREST-BGC use spatially explicit meteorological inputs from models such as MT-CLIM (Band et. al., 1991). Accurate estimates of temperature are critical to the performance of the above models. In addition to those involved in temperature modeling, temperature prediction at unsampled sites is of interest t o individuals involved in fire management, resource management, and spraying or seeding operations.

Accurate measurements of temperature are also of interests to scientists studying the "greenhouse effect" - global warming via the entrapment of longwave radiation due to certain gases such as carbon dioxide. While there is disagreement on the extent of global warming, most scientists estimate its effects between 0.50F to 1.00F (Handcock and Wallis, 1994). Clearly, even a small bias resulting from the interpolation method used would affect concl usions reached by scientists studying the greenhouse phenomena.

Accurate temperature estimates are critical in the calibration of satellite sensors. Satellite surface temperature retrieval in mountainous terrain is complicated by the high variability of occurring temperatures and complex terrain features. While sate llite surface temperature retrieval appears to be a promising technology, surface variations have been shown to bias temperature measurements upwards of 3.00C (Lipton, 1992). As bias is systematic, satellite deri ved temperature estimates calibrated with accurate ground truth, may offer cost effective temperature estimates where data are sparse.

Given a set of meteorological data, researchers are confronted with a variety of stochastic and deterministic interpolation methods to estimate meteorological variables at unsampled locations. Spatial interpolation is often an important first step in tak ing irregular point data and converting it for use in a GIS. Depending on the spatial attributes of the data, accuracies vary widely among different spatial interpolation methods (MacEachren and Davidson, 1987; and Rhind, 1975). The choice of spatial interpolator is especially important in mountainous regions where data collection are sparse and variables may change over short spatial scales.

While there have been comparisons of interpolation methods, few research efforts have been directed towards comparing the effectiveness of different spatial interpolators in predicting temperature (eg. Van Kuilenburg et. al. (1982); Dubrule (1983); Puente and Bras (1986); Bardossy et. al. (1987); Laslett et. al. (1987); and Phil lips et. al. (1992). A review of the literature indicates that a regionalized variable such as temperature, which is strongly correlated with elevation, would be well disposed to kriging and cokriging. Due to the additional effort kriging and cokrig ing entails, it was decided to compare the effectiveness of kriging and cokriging in estimating maximum and minimum temperature at unsampled locations with other less computationally intensive techniques such as inverse distance weighted averaging, cubic splining, the trend surface analysis (TSA), polynomial regression and lapse rate methods. Kriging and cokriging have also received some criticism due to the subjective nature of variogram fitting - a central component of kriging (P hillips and Watson, 1986). In addition to the aforementioned methods, this research introduces optimal inverse distance weighting where the inverse weighting parameter is chosen on the basis of minimum mean absolute error.

In this study, eight spatial interpolation techniques were compared across two regions (eastern and western North America), two temperature variates (tmax and tmin), and three temporal scales (10 year mean, seasonal mean, daily). Each interpolation techn ique was compared on the basis of bias, mean absolute error (MAE), and mean squared error (MSE). As the true temperature surface was not known, the comparison statistics were obtained using cross validation where one data point is withheld and the remain ing data points are used to predict at the withheld point. To obtain the cross validation statistics, the technique is repeated n times. In addition, the effects of data variance, data correlation with elevation, and lapse rate on MAE were investigated . Summary statistics were used to determine if any method was significantly better than the methods tested on the basis of bias, MAE, and MSE. Summary statistics were also used to determine whether the temperature variate (tmax or tmin) or temporal scal e (10 year mean, seasonal, or daily) affected interpolation.

2.0 Spatial Interpolation

Spatial interpolators may be used to estimate temperature at unsampled sites. Spatial interpolation can also be used when preparing irregularly scattered data to construct a contour map or contour surface, which is a two-dimensional representation of a t hree dimensional surface. All eight spatial interpolation methods investigated accept irregularly scattered data and can create a regular grid of interpolated points amenable to contouring. The spatial interpolation methods differ in their assumptions, local or global perspective, and deterministic or stochastic nature.

Inverse distance weighted averaging (IDWA) is a deterministic estimation method where values at unsampled points are determined by a linear combination of values at known sampled points. Distance-based weighting methods have been used to interpolate clim atic data (Legates and Willmont, 1990). IDWA makes the assumption that values closer to the unsampled location are more representative of the value to be estimated than samples further away. Weights change according to the l inear distance of the samples from the unsampled point. The spatial arrangement of the samples does not affect the weights. IDWA has seen extensive implementation in the mining industry due to its ease of use. IDWA has also been shown to work well with noisy data. The choice of power parameter in IDWA can significantly affect the interpolation results. As the power parameter increases, IDWA approaches the nearest neighbor interpolation method where the interpolated value simply takes on the value of the closest sample point. Optimal inverse distance weighting is a form of IDWA where the power parameter is chosen on the basis of minimum mean absolute error.

Splining is a deterministic technique to represent two dimensional curves on three dimensional surfaces (Eckstein, 1989; Hutchinson and Gessler, 1994). Splining may be thought of as the mathemat ical equivalent of fitting a long flexible ruler to a series of data points. Like its physical counterpart, the mathematical spline function is constrained at defined points. Splines assume smoothness of variation. Splines have the advantage of creatin g curves and contour lines which are visually appealing. Some of splining's disadvantages are that no estimates of error are given and that splining may mask uncertainty present in the data. Splines are typically used for creating contour lines from den se regularly-spaced data. Splining may, however, be used for interpolation of irregularly-spaced data.

Polynomial regression is a stochastic, global technique which fits the variable of interest to some linear combination of regressor variables (Myers, 1990). In this case the variable of interest being temperature and the regress or variables being the weather station's X, Y, and Z coordinates. Typically, the goal when using polynomial regression is to obtain the best fit with the simplest model. The addition of regressor variables which do not contribute significantly to the mo del has the unwanted effect of increasing multicollinearity. Multicollinearity may negatively affect the model's ability to predict outside the convex hull of data points (Myers, 1990). In this study, temperature was fitted to first, second, and third order polynomial models of the X and Y coordinates plus elevation. The model with lowest Mallow's Cp statistic was then chosen for predicting temperature.

Trend surface analysis (TSA) can be thought of as a subset of polynomial regression. TSA is a stochastic technique which separates the data into regional trends and local variations. The regional component of TSA can be thought of as a regression surfac e fit to the data, while the local variations can be thought of as a map of residuals. Values at unsampled locations may be estimated using the mathematical relationship between the locational variables X, Y and the regionalized meteorological variable o f interest. In this study, temperature was fitted to a third order polynomial. A third order polynomial was assumed to be sufficient to capture regional temperature variations.

TSA differs from polynomial regression above in that elevation is not used in estimating temperature and TSA uses all regressors variables, not a subset chosen on the basis of Mallow's Cp. Estimation using TSA is limited by problems associated with edge effects and multicollinearity caused by spatial autocorrelation. TSA assumes errors are independent. In addition to use as a spatial interpolation technique, TSA receives use in the removal of broad trends prior to further spatial analysis such as krigi ng.

The lapse rate method uses the relationship between temperature and elevation for a region to estimate temperatures at unsampled sites. Typically, temperatures decreases as elevation increases. This relationship between temperature and elevation is know n as the lapse rate. The lapse rate method uses the temperature value of the nearest weather station and the difference in elevation to estimate temperature at the unsampled site. To estimate temperature at an unsampled site, the difference in elevation is multiplied by the lapse rate and the subsequent number is added to or subtracted from the weather station temperature to yield the site temperature. The lapse rate method makes the assumption that the lapse rate is constant for the study region.

Kriging is a stochastic technique similar to inverse distance weighted averaging in that it uses a linear combination of weights at known points to estimate the value at an unknown point. Kriging is named after D.L. Krige, who used kriging's underlying th eory to estimate ore content. The general formula of kriging however, was developed by Matheron (1969). Kriging uses a semivariogram, a measure of spatial correlation between two points, so the weights change according to th e spatial arrangement of the samples. Unlike other estimation procedures investigated, kriging provides a measure of the error or uncertainty of the estimated surface. In addition, kriging will not produce edge-effects resulting from trying to force a p olynomial to fit the data as with TSA.

Cokriging is similar to kriging except it uses additional covariates, usually more intensely sampled, to assist in prediction. Cokriging is most effective when the covariates are highly correlated. Both kriging and cokriging assume homogeneity of first differences. While kriging is considered the best linear unbiased spatial predictor (BLUP), there are problems of nonstationarity in real-world data sets.

3.0 Research Methodology

This comparative analysis used National Weather Service (EarthInfo, 1992) and SNOTEL temperature data, and USGS 3 arc-second digital elevation data. Two test regions were used. The regions were selected for their contrasting densities of National Weath er Service (NWS) and SNOTEL weather monitoring stations. Region 1 has stations well dispersed throughout the study area while Region 2 has stations clustered around population centers. In addition, in Region 1, the station elevations are representative of the region elevations. In Region 2, higher elevations are under represented by weather monitoring stations. Region 1 is located in the upper piedmont region of the eastern United States at 34 to 37 degrees North latitude and 80 to 85 degrees West lon gitude. The region encompasses portion of the states of Virginia, North Carolina, South Carolina, and Georgia. In Region 1 there are 146 NWS stations.

The NWS data were read from the EarthInfo CD-ROM and then filtered using dBaseIV software. The SNOTEL data were downloaded via modem access and converted from their ASCII file format into dBaseIV format. This conversion was necessary for the filtering a nd creation of the data sets used in this comparative analysis. SNOTEL data are in a different format from NWS data. In addition, SNOTEL data are record in degrees Celsius where NWS data are recorded in degrees Fahrenheit. All SNOTEL temperature readin gs were converted to degrees Fahrenheit and reformatted to conform to the NWS data format.

The NWS data is known to have erroneous daily values resulting from errors in data-entry, data-recording, and data-formatting errors (Reek, Doty, and Owen, 1992). In their paper, Reek et. al. propose a deterministic approach for the removal of these systematic errors. Since 1991, NCDC has implemented the approach proposed by Reek et. al. The pilot program entitled Validation of Historical Daily Data (ValHiDD) was designed to improve upon an older quality assurance program calle d Geographical Edit and Analysis (GEA) NCDC estimates that errors in its historical data are small, in the area of .05% or less. In addition, most errors associated with NWS data are associated with older (pre 1960) temperature readings. As this resear ch only deals with recent (1980 - 1990) data, the data can be assumed to be relatively error free. In addition, since all interpolation methods used the same 60 data sets, any errors in the data will apply across all interpolation methods. Hence, the co nclusions of a comparative analysis would not be affected.

Missing data were handled differently depending on the temporal scale. For 10 year means, missing monthly means were averaged from adjacent months. A given weather station could have as many as two months missing out of any given year and still be inclu ded in the 10 year mean. In cases where monthly averages were missing, the first recourse of action was to attempt to calculate the monthly average from daily data. If 20 or more days were not missing for that given month, the average would be calculate d and used as the monthly average. Typically, certain stations were prone to missing data. This data averaging was necessary to ensure an adequate number of stations for the 10 year mean. Even with data averaging, nearly one third of candidate stations within the test region were dropped from this analysis due to missing data. For seasonal means, missing monthly averages were calculated from daily averages when there were 20 or more daily temperature available. For daily temperature values, stations with missing data were excluded from the analysis.

Each interpolation technique was compared on the basis of bias, mean absolute error (MAE), and mean squared error (MSE). For kriging and cokriging, cross validation techniques are used to choose the best semivariogram model from among candidate models (s pherical, exponential, or Guassian). In addition, cross validation techniques are used to select the search radius which minimize the kriging variance.

For kriging and cokriging, semivariogram modeling was accomplished using Geo-EAS software (Englund, 1988) and VARIOWIN software (Pannatier, 1994). Kriging and Cokriging are accomplished using t he GSLIB software library (Deutsch, 1992). Finally, the results of each interpolation method are examined visually using Sigma PlotTM.

This comparative effort studied three temporal scales: 10 year average, seasonal, and daily. A total of 408 temperature contour surfaces are generated. The 408 surfaces result from 10 year mean, maximum, and minimum temperatures for two regions for six interpolation techniques (1*2*2*6 = 24); four seasonal minimum and maximum temperatures for two regions for six interpolation techniques (4*2*2*6 = 96); and 12 daily mean, maximum, and minimum temperatures for two regions for six interpolation techniques (12*2*2*6 = 288). The results of the comparative analysis are presented in tabular as well as graphical format.

Summary statistics rather than analysis of variance (ANOVA) were used to determine if any interpolation method was significantly better than the methods tested on the basis of bias, MAE, and MSE. The reason for this choice was that hypothesis testing usi ng ANOVA assumes that the means compared are drawn from populations with a common variance. As this analysis provides a comparison across different temporal scales and geographics regions, the assumption of a common variance is not valid. Data attribute s were also investigated to determine whether data variance, data correlation with elevation, or lapse rate significantly affect interpolation. Because these spatial metrics are easy to calculate, they can be determined prior to interpolation to determin e which method may be most appropriate.

4.0 Results

The analysis found evidence that certain apriori data characteristics influence the choice of spatial interpolation technique. Temperature range, temperature variance, and temperature correlation with elevation all influence the choice of interpolation t echnique. Spatial scale also impacts interpolation. In addition, the relative spatial density and distribution of sampling stations may influence the choice of interpolation technique. These conclusions concur with MacEachren and Davidson (1987) who concluded that data measurement accuracy, data density, data distribution, and spatial variability had the greatest influence on interpolation accuracy.

4.1 Effect of Data Attributes on Temperature Interpolation

Overall, polynomial regression was most representative of the original data and had the lowest MAE value of methods ranked. Higher correlations between elevation and temperature favored polynomial regression and the lapse rate method. Inverse distance s quared, optimal inverse distance, and kriging showed a similar robustness to apriori data range, correlation (between elevation and temperature), and variance. Of all methods studied, splining seemed to be most sensitive to apriori data characteristics. TSA did not produce results representative of the original data. TSA's interpolated temperature range was typically more narrow than the original data. Kriging was favored over optimal inverse distance when data were anisotropic. When data were isotro pic, optimal inverse distance averaging (ODA) was favored. When data variance was high or correlation between temperature and elevation were low, cokriging had specking or "birds eye" effects around station locations. Overall, cokriging gave visually im plausible results. The fact that a single interpolation method, polynomial regression, showed clear superiority over other methods based on MAE is significant. Robeson (1993), who studied various spatial interpolator effectiv eness in predicting global temperature, found cross validation errors were approximately the same for all interpolation methods. Robeson, however, did not compare polynomial regression or methods which used elevation as ancillary information in estimatin g temperature. The differences among those interpolation methods which did not use elevation as ancillary information were not all that significant.

For all cases tested, kriging had lower MAE values when the data were anisotropic. When the data were isotropic, optimal inverse distance performed better than kriging based on MAE. Large temperature variances and temperature ranges tended to increase i nterpolator MAE. Higher correlations between temperature and elevation tended to favor polynomial regression over other interpolation techniques. For 10 year temperature means, the only situation where polynomial regression was not ranked highest based on MAE, were when correlations between elevation and temperature were lower than 0.72. For seasonal and daily data, the choice of spatial interpolator became less clear when correlations between elevation and temperature fell below 0.60. Above correlati ons of 0.60, regression appears to be the interpolator of choice.

The effects of landscape complexity did not directly affect the choice of spatial interpolator. Data attributes, however, change with landscape complexity. Region 2 had greater landscape complexity than Region 1. As a result of this increased landscape complexity, Region 2 had greater temperature variances and observed temperature ranges across all temporal scales.

In general, the results indicate that increased variance and data range result in decreased interpolator accuracy as indicated by higher MAE values. The results also indicate that interpolation techniques which use ancillary elevation information to pred ict temperature benefit from higher correlations between elevation and temperature. Of data attributes investigated, correlation between elevation and temperature and data temperature variance had strong influences of predictor performance

4.2 Effect of Temporal Scale on Temperature Interpolation

Temporal scale affects the choice of spatial interpolator as temperature range, temperature variance, and temperature correlation with elevation, all change with temporal scale. As one moves from ten year means to seasonal to daily temperatures, temperat ure variances tend to increase, temperature ranges are greater, and correlation between temperature and elevation was far more variable. There was a general increase in temperature range as one progresses from 10 year means to seasonal means to daily tem peratures. There was also a general increase in temperature variability as temporal scale was reduced. Correlations between elevation and temperature become more variable as temporal scale was reduced. This may be due to temperature inversions for a gi ven day. Such inversions may be smoothed when temperature data are viewed in aggregate.

Range, variance, and correlation are important attributes to consider when selecting a spatial interpolator. Where temporal scales are short, preliminary data analyses are especially important to determine the suitability of a particular interpolation te chnique. The lapse rate method, cokriging, and polynomial regression cannot be recommended when correlations between temperature and elevation are below 0.72. When temperature variances are large, the performance of all interpolation techniques suffers. The larger MAE values of daily minimum and maximum temperature interpolation can be attributed to higher temperature ranges and variances.

Results from the interpolation of ten year mean maximum and minimum temperature shows that MAE increases with increasing temperature variance and temperature range. Ten year mean results indicate lower MAE values across all interpolation techniques as th e correlation between temperature and elevation increased. Polynomial regression was ranked first based on MAE for all cases except for Region 2 minimum temperature, where the correlation between elevation and temperature was 0.71. Region 2 minimum temp erature, which had the highest data variance and range also had the poorest interpolator performance based on MAE.

Results from the interpolation of seasonal mean maximum and minimum temperature was not as clear as for 10 year means. The results for seasonal means indicate that trend surface analysis (TSA), inverse distance squares (IDA), optimal inverse distance (OD A), and kriging were all rather robust to the effects of temperature range, temperature variance, and temperature correlation with elevation. The seasonal results did indicate much lower MAE values for polynomial regression across all seasons. In additi on, there was clear evidence that polynomial regression gives more accurate results which are representative of the original range of the data when correlations between temperature and elevation are high. These trends are similar across both Region 1 and Region 2. Region 2 had a wider range of temperature values and a higher overall variance than Region 1. Increased correlations between elevation and temperature resulted in better polynomial regression performance for Region 2.

Results from the interpolation of Region 1 daily maximum and minimum temperature indicates that increasing temperature variance affects interpolator performance negatively. Increased correlations between elevation and temperature had a positive effect on interpolator performance. Temperature range did not seem to affect interpolator performance. Results for Region 2 daily maximum and minimum temperature differed from Region 1 results in that temperature range had a negative effect on interpolator pe rformance. As the temperature range increased, MAE values across all interpolators increased significantly. The effects of temperature variance and correlation between elevation and temperature were the same as for Region 1. As temperature variance inc reased, MAE values across all interpolators increased. As correlations between elevation and temperature increases, MAE values dropped significantly for those interpolation methods which used elevation as ancillary information.

4.3 Performance Considerations by Technique

4.3.1 Inverse distance squared

Inverse distance squared (IDA) gave consistent, albeit, poor performance across all temporal and regional scales. Where data are sparse, as in Region 2, IDA's results were implausible. IDA also suffered from discontinuities at station locations resultin g in temperature peaks. These temperature "peaks" resulted in "birds eye" patterns in the contour maps. IDA's one advantage was that it consistently adhered to the original temperature range of the data. Where the data are representative of the surface being interpolated, as in Region 1, this is advantageous. Where the data are not representative of the surface being modeled, as in Region 2, it may result in interpolation biases. ODA is recommended over IDA because ODA will always yield equal or bett er MAE results than IDA, since the power parameter is chosen on the basis of minimum MAE.

4.3.2 Optimal inverse distance

Optimal inverse distance (ODA) had consistently better results than IDA or kriging. By design, it's power parameter was chosen on the basis on minimum MAE. Therefore, it was equal to or better than inverse distance squared. When the data were isotropic , IDA had lower MAE values than kriging. ODA, however, was not always as visually plausible as kriging. When the data are not correlated and isotropic, the most preferred method appears to be ODA. Kriging appears to be more visually plausible than ODA, but the summary statistics indicate that it is not always representative of the original data's range. Given ODA's ease of use and superiority over kriging when data are isotropic, it is recommended in cases where ancillary information correlated with te mperature is not available.

4.3.3 Trend surface analysis

Overall, TSA can not be recommended for temperature interpolation. TSA was not representative of the original data range. TSA tended to capture broad regional trends, but due to bias introduced by multicollinearity, these trends are suspect. In Region 2, where station distribution resulted in some extrapolation beyond the convex data hull, TSA had interpolated temperatures well beyond the original data range. Tabios and Salas (1985), who compared kriging, optimal inverse distance, and splining, also c oncluded that TSA gave the poorest results of methods compared while kriging and ODA gave the more accurate results in precipitation estimation.

4.3.4 Polynomial regression

Based on MAE, the regression method was clearly superior to all other methods ranked. For correlations between elevation and temperature above .72 and for variances less than 30F2, regression was preferred over the other spatial interpolation methods. W here correlations between temperature and elevation were very low (under 0.20) optimal inverse distance and kriging had lower MAE values. Polynomial regression performance did not appear to be affected by data range. In this study, temperature was fitte d to first, second, and third order polynomials of the X and Y coordinates plus elevation. The model with lowest Mallow's Cp statistic was then chosen for predicting temperature.

$$Temp = b + b_1X + b_2Y + b_3X2 + b_4Y2 + b_5XY + b_6X3 + b_7Y3 + b_8X2Y + b_9XY2 + b_{10}Z$$

When correlations between temperature and elevation are low, the elevation term, Z, drops out and what is left is similar to TSA. When the elevation term of polynomial regression drops out, the method is not recommended. Of the methods tested, polynomia l regression gave the most visually plausible results. When elevation is highly correlated with temperature, the temperature contours tend to follow elevation contours. In addition, broad regional trends are captured by the x and y variables when they a re significant. Care must be taken with polynomial regression to ensure the results are representative of the original data range. Where station elevations are not representative of regional elevations (such as in Region 2) care must be taken in compari ng observed and interpolated data. For Region 2, temperature contours which used elevation were assumed to have slightly lower mean temperatures with somewhat lower minimum temperatures than the observed data because the DEM contained higher elevations t han did the stations.

4.3.5 Cubic splining

While similar to kriging mathematically, splining gave poor visual and cross validation results. Where data variances were high, splining tended to have interpolated values well outside the observed data range. For lower correlations splining had very h igh MAE values. In general, splining had high MAE values across all temporal scales, across both regions, for both minimum and maximum temperatures. Cubic splining was especially poor as an interpolator in Region 2. Cubic splining is generally not reco mmended for interpolation of irregularly-spaced data (Eckstein, 1989; Hutchinson and Gessler, 1994). Dubrule (1983) compared splines and kriging and concluded that sp lines are useful for quickly obtaining a clear map showing the main feature of the variable, but it is not an accurate interpolator. Dubrule's results showed splining had more outliers than kriging and that kriging had lower MAE values. This study suppo rts Dubrule's conclusions.

4.3.6 Kriging

Of the methods which did not use elevation as ancillary information, kriging was most visually plausible. Kriging gave better results than optimal inverse distance when data were anisotropic. In some cases for Region 1 data, kriging appeared similar to splining, but had lower MAE values than splining for every case tested. Kriging typically had lower MAE values than inverse distance squared. Ishida and Kawashima (1993) also found kriging estimates to be better than inverse d istance squared. Using available semivariogram modeling tools (Pannatier 1994), one can determine when data are anisotropic, select the best fitting semivariogram model, and obtain kriged results easily. Given the added steps kriging entails, and the ma rginal benefits of kriging over ODA, the justification of kriging over ODA is questionable. Perhaps the greatest advantage of kriging is that the geostatistical process provides the users greater information about the spatial variability of the regionali zed variable of interest via semivariograms and variogram surfaces.

4.3.7 Cokriging

Cokriging's contour plots had circular patterns centered on station locations in Region 1 and unlikely tessellation patterns in Region 2. When data variance is low and correlation between temperature and elevation are high, Cokriging appears to have les s specking or "birds eye" effects When elevation and temperature are not correlated cokriging bears a strong resemblance to kriging. This resemblance is to be expected as the elevation component of cokriging is not significant. While cross validation s tatistics for cokriging were not computed, visual inspection of the cokriged temperature surface showed cokriging followed elevation contours more closely than did kriging when correlations between elevation and temperature were significant. These result s correspond with Cob (1990) who found cokriging more closely reflected the elevation features of the climatic regions when using elevation as a covariate. Cob also found cokriging to be more accurate than kriging in predicting pr ecipitation when elevation correlations were significant. While unproven in this study, one would expect cokriging to be more accurate than kriging in predicting temperature when correlations between elevation and temperature are significant.

In general, cokriging did not perform well in this study. The restrictions placed on parameter selection due to semivariogram models being strictly positive definite may have adversely affected cokriging results. In addition, the high variability of tem perature data does not appear to lend itself to cokriging. The greater the data variability, the more difficult the data are to fit. Cokriging requires the fitting of two semivariograms and one semi-crossvariogram for each data set.

4.3.8 Lapse rate method

While the lapse rate method performed poorly in terms of MAE, its results were more plausible than methods which did not use elevation as ancillary information. This was especially true where station elevations were not representative of regional elevati ons, such as in Region 2. Lapse methods was preferable over cokriging on the basis of visual plausibility and adherence to the original data range. While preferable over cokriging, the lapse rate method showed some banding effects and island-like isothe rmal tessellations around certain influential stations. Outlier stations were less noticeable with polynomial regression than with the lapse rate method. When elevation and temperature are not correlated, then the lapse rate method degrades into a neare st neighbor method where interpolated values simply take on the value of the nearest station point.

5.0 Recommendation for Further Research

Given the effectiveness of the polynomial regression method, it would be interesting to use other intensely sampled information highly correlated with temperature. From the DEM one can obtain slope and aspect information which could be used as additional covariates provided this slope and aspect information could be accurately determined for all station locations. The use of the 6th thermal band in Landsat satellite data or other remotely sensed information may be useful in temperature estimation. Running et al. (1987) proposes the augmentation of climate models with data from meteorological satellites to improve surface climate estimates. Lipton (1992) found that slope and aspect variations may b ias satellite surface temperature retrievals. The effects of such bias would have to be ameliorated prior to the use of satellite information for temperature estimation.

Additional data would enhance this research effort. The application of this research methodology to larger data sets such as the NCAR World Monthly Surface Station Climatology (Spangler and Jenne, 1988) would be of interest as no study has been conducted using elevation to assist in temperature prediction at global scales.

References

- Band et. al., 1991, Forest ecosystem processes at the watershed scale: basis for distributed simulation , Ecological Modelling, Vol. 56, p171-196.

- Bardossy, A. I. Bogardi, and L. Duckstein, 1987, "A geostatistical model of reservoir deposition", Water Resources Research, 23, p510-514.
- Boyer, D.G., 1984, Estimation of Daily Temperature Means Using Elevation and Latitude in Mountainous Terrain, Water Resources Bulletin, no. 4, p583-588.
- Cob, Antonio Martinez, 1990, Multivariate Geostatistical Analysis of Evapotranspiration and Elevation for Various Climatic Regimes in Oregon, Doctoral Disseratation.
- Cressie, Noel A.C., 1993, Statistics for Spatial Data, John Wiley & Sons, Inc., New York, 900p. [revised edition]
- Creutin, J.D. and C. Obled, 1982, "Objective analysis and mapping techniques for rainfall fields: An objective comparison", Water Resources Research, 18, p413-431.
- Deutsch, C.V., and A.G. Journel, 1992, GSLIB: Geostatistical software library and user's guide, Oxford University Press, New York, 340p.
- Dubrule, Oliver, 1983, Two methods with different objectives: Splines and kriging, Mathematical Geology, 15, No. 2, p245-257.
- Eckstein, B.A., 1989, Evaluation of spline and weighted average interpolation algorithms, Computers & Geosciences, 15, No. 1, p79-94.
- Handcock, M.S., and J.R. Wallis, 1994, An Approach to Statistical Spatial-Temporal Modeling of Meterological Fields, JASA, Vol. 84, No. 426, p368-378.
- Hutchinson, M.F., and P.E. Gessler, 1994, Splines - more than just a smooth interpolator, Geoderma, Vol. 62, p45-67.
- Ishida, T. and S. Kawashima, 1993, Use of Cokriging to Estimate Surface Air Temperature from Elevation, Theor. Appl. Climatol., vol. 47, p147-157.
- Laslett, G.M, A.B. McBratney, P.J. Pahl, and M.F. Hutchinson, 1987, "Comparison of several spatial prediction methods for soil pH", Journal of Soil Science, 38, p325-341.
- Legates, D.R. and C.J. Willmott, 1990, Mean Seasonal and Spatial Variability in Global Surface Air Temperature, Theor. Appl. Climatol., vol. 41, p11-21.
- Lipton, A.E., 1992, Effects of Slope and Aspect Variations on Satellite Surface Temperature Retrivals and Mesoscale Analysis in Mountainous Terrain, Journal of Applied Meterology, vol. 31, no. 3, p255-264.
- MacEachren, A.M., J.V. Davidson, Sampling and Isometric Mapping of Continuous Geographic Surfaces, The American Cartographer, Vol. 14, No. 4, 1987, p299-320.

Matheron, G., 1963, Principles of geostatistics, Economic Geology, 58, p. 1246-1266.

- Myers, R.H., 1990, "Classical and Modern Regression with Applications", PWS-Kent Publishing, Boston, MA, p488.

- Olea, Ricardo A., 1975, Optimum mapping techniques using regionalized variable theory Kansas Geological Survey Series on Spatial Analysis No. 2,137p.

- Philip, G.M. and D.F. Watson, 1986, Matheronian Geostatistics - Quo Vadis?, Mathematical Geology, 18, No. 1, p93-117.

- Phillips, D.L., J. Dolph, and D. Marks, 1992, "A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain", Agricultural and Forest Meterology, 58, p119-141.

- Puente, Carlos E., and Rafael L. Bras, 1986, Disjunctive kriging, universal kriging, or no kriging: Small sample results with simulated fields, Mathematical Geology, 18, No. 3, p287-305.

- Reek, T., R. Doty, and T.W. Owen, 1992, A Deterministic Approach to the Validation of Historical Daily Temperature and Precipitation Data from the Cooperative Network, Bulletin American Meterorological Society.

- Rhind, D., 1975, A skeletal overview of spatial interpolation techniques, Computer Applications, Vol. 2, No. 3/4, p293-309.

- Robeson, S.M.,1993, Spatial Interpolation, Network Bias, and Terrestrial Air Temperature Variability, Publications in Climatology, C.W. Thornthwaite Associates Laboratory of Climatology,p51.

- Running S.W., R.R. Nemaini, R.D. Hungerford, 1987, Extrapolation of synoptic meteorolgoical data in mountainous terrain and its use simulating forest evapotranspiration rate and photosynthesis, Canadian Journal of Forest Res earch, Vol. 17, p472 - 483.

- Spangler, W. M. L. and R.L. Jenne, 1988, "World Monthly Surface Station Climatology", NCAR, Scientific Computing Division, Boulder, Colorado.

- Van Kuilenburg, J., J.J. De Gruijter, B.A. Marsman, and J. Bouma, 1982, Accuracy of spatial interpolation between point data on soil moisture supply capacity, compared with estimates from mapping units, Geoderma, 27 , p311-325.

Smart Interpolation of Climate Variables

Cort J. Willmott
University of Delaware

Modern spatial interpolation procedures rely primarily on geometric and statistical ("geostatistical") relationships among observations. Interpolation accuracy then is constrained by the extent to which available observations represent the variable of interest, as well as by the inherent ability of geostatistical constructs to estimate missing observations. Smart (SiMple but AccuRaTe) interplators differ from their geostatistical counterparts in that they are constructed primarily from knowledge or understanding of the processes which produced the spatial variation within the field of interest. Their mathematical form, in other words, arises mainly from a conceptualization of process, rather than from statistical considerations. Geostatistical interpolation may be embedded within a smart interpolator, but in a subordinate role. Smart interpolators developed by the author and his graduate students (for interpolating average air temperature and precipitation) are used to illustrate smart approaches. Evaluation of interpolation error is an important aspect of the interpolator building process; therefore, a simple cross-validation approach for estimating interpolation error fields also is presented. Measures of the average cross-validation error or difference, based on the mean-absolute error (MAE), are recommended as well.

Barbara P. Buttenfield and Mark P. Kumler

Tools for Browsing Environmental Data: The Alexandria Digital Library Interface

To service those who need digital data, new products appear with increasing frequency, and one can access increasing quantities of geographic data on the Internet. Paradoxically, as more data become available they become more difficult to locate, to download, and to certify as valid. A major challenge in the coming decade is to enhance the accessibility, communication and use of geographically referenced data. The Alexandria Digital Library Project implements a software testbed delivering comprehensive library services to browse and retrieve maps, imagery, historical air photos, and other georeferenced digital data distributed on local and wide-area (Internet) networks.

A working prototype of the Library is complete. User evaluation plays an important role in testing the effectiveness of current software functions for browsing environmental data. The current interface design embeds online user evaluation mechanisms, including object oriented interactive logging to monitor use patterns and use error patterns. Interactive dialog tools enable users to annotate specific system commands and behavior that delight or confuse them. Logs and user dialog are analyzed to guide interface refinement. The intention is to optimize an interface for browsing environmental data on the Internet. The interface and user evaluation tools will be demonstrated at the conference.

CHANGING NEEDS FOR BROWSING ENVIRONMENTAL DATA

To service those who need digital data, new products appear with increasing frequency, and one can access increasing quantities of geographic data on the Internet. Federal agencies that produce and distribute environmental datasets are converting from physical to electronic distribution mechanisms. Data enhancement is increasingly outsourced to private companies who add value to federal products, repackage and redistribute them on the Internet. Environmental scientists who previously placed orders for data on magnetic tape or CD-ROM from agencies or companies can now gain access via the Internet. The challenge for the scientist is to navigate the ever-increasing volume of information and to locate and download appropriate data. This requires a new set of skills for the environmental scientist, and also requires provision of new tools for generalized and specialized data delivery. A major challenge for the GIS community in the coming decade is to enhance the accessibility to geographically referenced digital environmental data.

This paper describes a continuing research project directed at this data accessibility challenge. The Alexandria Digital Library Project implements a software testbed delivering comprehensive library services to browse and retrieve maps, imagery, historical air photos, and other georeferenced digital data distributed on local and wide-area (Internet) networks.

The primary project site is at the University of California - Santa Barbara, with a second site for user evaluation research at the University of Colorado. One unique aspect of the Alexandria Project is that system implementation efforts are informed by user feedback and evaluation. That is, attention is simultaneously directed to implementing software tools and to evaluating information requirements and skills of various types of environmental data users. This complicates project management considerably, however it also provides an important reality check to insure that system design at every stage of development meets most user requirements. The strategy also provides an opportunity to evaluate the user evaluation mechanism, streamlining and minimizing intrusive requests for user feedback. The project is extensive, and currently under construction. The paper will overview the system as a whole, focusing in particular upon interface design and evaluation.

ALEXANDRIA DIGITAL LIBRARY SERVICES

The project will provide comprehensive services for collections including digitized maps and images, spatially referenced digital data, and other environmental information (historical air photographs, sets of feature coordinates, and metadata descriptions). The intention is that data collections will be distributed and delivered across wide-area networks, allowing access to the Library via the Internet.

Users can browse Library holdings electronically and search by spatial or temporal location or by metadata content. Spatial searches by placename or by spatial footprint can be refined according to specific time periods, data resolution, data category (satellite image, topographic map, geologic map, etc.). Efforts are underway to implement browsing tools based on collections maintenance criteria (map sheets having multiple editions, e.g.) and based on information content. A hypothetical example of information-based browsing would be a user request to find a map covering the driftless region of Wisconsin and containing geologic features, to display the map and any air photos of the same footprint shot prior to 1940, and to provide citations to technical books describing the possible glaciation of this region.

The project has proceeded in phases, and multiple versions of the Library are currently under development. Reasons for this include the need to have a stable Library system in place to support user evaluation studies, the need to demonstrate that individual software modules are operational before they are added to the general testbed, and to provide system designers a platform for experimentation and benchmarking. The first phase produced a rapid prototype running commercial off-the-shelf software (ArcView) on a UNIX platform. The rapid prototype was completed in Spring, 1995, and has served since that time as the major platform for user interface evaluation efforts (described below). A subset of the rapid prototype has been ported to a Windows platform, and burned onto CD-ROM. The current phase extends rapid prototype functions in a World-Wide Web environment. For example, users may custom tailor the look of the query interface in this version. The Web testbed has nonetheless presented major challenges for system designers, given the limited graphics functions currently available on the Web. For example, no Web browsers currently available provide "lasso-ing" functions, which form a basis for user-defined footprint selection. The Web testbed is currently operational, but is not yet completely stable. It is expected to become publicly accessible later this spring. A homepage announcing its availability is provided at the end of this paper.

ALEXANDRIA SYSTEM ARCHITECTURE

System architecture includes a storage component, a catalog component, an ingest component, and an interface component. The storage component is designed to accommodate very large collections of very large digital objects. Environmental data is alternatively characterized by high resolution multispectral raster data, and overlaid themes of vector data compiled at multiple map scales. Storage requirements are large. For example, an analog air photograph scanned at 600 dots-per-inch commonly requires 30 MB (90 MB for color) per archived image. A single collection of historical photography containing hundreds or thousands of images could require storage on the order of single terabytes at the point of archival. Distributed storage provides the only feasible architecture for multiple datasets, and Internet protocols (e.g., Z39.50) are being implemented to handle delivery and transfers. Current system holdings focus on the southern California region.

The catalog component is a special emphasis for current system development efforts. The catalog systematizes all types of information by which the Library holdings may be organized. By implication, the catalog contents form the basis for user browsing. An archive may only be searched on the items which are organized in its catalog. (One reason the Web is difficult to navigate is that it lacks a catalog.) The Alexandria catalog allows browsing by placename, by data them, by location (spatial footprint), by time (date of compilation), or by metadata as defined by FGDC/USMARC standards. Placenames are provided by the Geographic Names Information System (GNIS) gazetteer, which includes 1.8 million names of US features in 15 classes, and by the Board of Geographic Names (BGN) gazetteer, including 4.5 million names of land and undersea features. The catalog is stored in a central relational database (Sybase) housed in Santa Barbara. Metadata records are stored similarly, using Microsoft Access. Extensions to the catalog will be content-based, to provide user capabilities for content-based browsing described above.

The ingest component currently provides for input of data, metadata, and catalog information. One should expect that eventually, users wishing to augment the Alexandria holdings will utilize ingest functions. Data ingest is accomplished by scanning analog material, by transfer of created metadata records from Microsoft Access, or from other sources (e.g., 450K frame-level records for NASA air photo database, 350K sheet-level records for map series (Geodex), and 100K USMARC map records from MELVYL). New catalog records must be created to catalog pointers to Web sites for digital spatial data, and for example to record metadata for air photography for four California counties.

The interface component is most visible to users. To some, there may appear to be no difference between the interface and the Library. Interface functions include tools for indexing, retrieval, and data access (browsing), tools to formulate queries by location, time, metadata, and (eventually) content. Interface utilities to guide image fusion, compression, and filtering may be used to override system defaults for data delivery and exploration. Interface display tools allow users to draw spatial footprints on a search map, for example.

All the interface functions will be evaluated. Since functions do not operate independently, the interface evaluation and re-design is a highly circular process. Low level functions are relatively easy to test, and include screen icon design and system command driven by keyboard and mouse. Many of these functions may be tested before they are embedded in the system. For example, one early experiment tested the amount of zoom provided by incremental user commands. Too-little zoom makes users impatient, and too-much startles them. Some users have requested user-specified variable zoom levels, although in the current Web implementation, this is not possible. Higher-level functions are more difficult to evaluate, and are summarized under the rubric of user satisfaction, discussed below.

USER INTERFACE EVALUATION

The following set of user requirements guided design of the rapid prototype:

* Range of skill levels for all types of users

* Graphical interface accessing multiple resolutions

* No manual required

* Ability to search on multiple data types
text, scanned imagery, map indexes, digital attributes, metadata

* Flexible search and query
intelligent georeferencing, object based query, metadata query

The goals of the User Evaluation team are threefold: first to evaluate the interface empirically and to provide feedback to the system designers; second to identify and respond to user requirements for the interface; and third, to research the application of interactive methods to interface evaluation. Accomplishment of these goals involves working with users, who include earth and space scientists, professional librarians, spatial data archivists, educators and students at all levels, government representatives at all levels, and Alexandria system designers.

Types of Collected User Data

Several types of data are collected to evaluate low-level and high-level interface functions. First, users are asked about previous experience using library collections, about their frequency of computer use, and whether they have access to online data catalogs or online services. This information develops a user profile, which may help to distinguish classes of users. The user profile which has developed after evaluating roughly 70 users is relatively homogeneous. Library use is frequent, although few of those tested (save for special collections librarians) are familiar with library special collections (e.g., map libraries). Almost all users are familiar with geographic and environmental digital data, and all are computer literate. Two aspects of the user profile distinguish between user classes, and these two both isolate students from all other user groups. Students are uniformly familiar with Internet use, while other user groups are split (roughly two-thirds are familiar). Conversely, students rarely work with online services, including online catalogs (e.g., MELVYL) and few subscribe to commercial access services (e.g.., America-Online, Prodigy) while other groups of users are split about half and half.

Second, both the rapid prototype and the Web versions have embedded within them capabilities for interactive transaction logging, to monitor use and use error patterns, and identify parts of the interface which need refinement. A transaction log records the sequence of menu buttons and tools that the user invokes, along with an anonymous user identifier and a timestamp. In some user logs, the transaction sequence will oscillate between one command and a second, indicating confusion. At least one menu button has been redesigned as a result of several transaction log oscillations, and a second interface function has been streamlined. A portion of a transaction log for user #0506-9438 is given below, monitoring a query and retrieval of a geologic map of California followed by a series of zoom and pan operations.

User 0506-9438 -- Transaction Log for Session 208
208, 1995/10/17, 09:02:32, 0506-9438, Tcl/Tk, QueryForm, QInput,
Submit=ALL;ALL;NULL;NULL;-122.5;-117.1;36.4;32.4
208, 1995/10/17, 09:05:48, 0506-9438, Tcl/Tk, QueryForm, Q-Input,
Submit=ALL;geologic maps;NULL;NULL;-180;180;90;-90
208, 1995/10/17, 09:06:25, 0506-9438, Tcl/Tk, QueryForm, Q-Input,
Submit=ALL;geologic map;NULL;NULL;-180;180;90;-90
208, 1995/10/17, 09:06:46, 0506-9438, Select, Retrieve Records, QueryEng,
2 hits; Retrieve records? true
208, 1995/10/17, 09:08:08, 0506-9438, Select, Selection Pad, MakeRecord,
#2; Geologic map California (southern half only)
208, 1995/10/17, 09:08:58, 0506-9438, Butt, Selection Pad, Bad,
Bad Idea -- This is not so great.
208, 1995/10/17, 09:09:02, 0506-9438, Butt, Selection Pad, comments,
Click here to tell us what is good or bad (or why)...
208, 1995/10/17, 09:10:17, 0506-9438, Note, Selection Pad, comments,
There's no way to know if what you clicked on has actually been accepted
or not - there's no hour glass prompt or something similar to let you know.
208, 1995/10/17, 08:48:14, 0506-9438, Tool, Search Map, ClickZoom,
Zoom into the area of interest
208, 1995/10/17, 08:48:16, 0506-9438, Tool, Search Map, ClickZoom,
Zoom into the area of interest
208, 1995/10/17, 08:48:30, 0506-9438, Tool, Search Map, ClickZoom,
Zoom into the area of interest
208, 1995/10/17, 08:49:07, 0506-9438, Butt, Search Map, good,
This is great!
208, 1995/10/17, 08:49:22, 0506-9438, Tool, Search Map, ClickZoom,
Zoom into the area of interest
208, 1995/10/17, 08:50:14, 0506-9438, Butt, Search Map, comments,
Click here to tell us what is good or bad (or why)...
208, 1995/10/17, 08:51:15, 0506-9438, Note, Search Map, comments,
Doesn't zoom in enough each interval; Is there a way to adjust how much
it zooms in?
208, 1995/10/17, 08:54:01, 0506-9438, Butt, Search Map, Unzoom,
Zoom back to previous scale

A third type of data is collected to evaluate user satisfaction, which was identified earlier as a high-level user function. Measures of satisfaction are associated with the user getting the information as requested. This can relate to the length of time the system takes to respond to a query, or to general vagaries of system behavior. Menu tools in the rapid prototype include three buttons by which the user may annotate a session, including a "good" button, a "bad" button, and a "notepad". In the transaction log above, the user activates the "bad" button and then writes a note asking that the system distinguish when it is working on a query as opposed to being hung. The user later activates the "good" button and notes a comment about variable levels of zooming in on a map. Not all users invoke the notepad tool, however, many utilize the "good" and "bad" buttons. These buttons identify which types of system activities invoke positive and negative responses, and their position in the transaction log identify where in the sequence of a particular task an interface becomes important to the user.

Summarizing the levels of satisfaction from the roughly 70 users to date, users overall feel that the Rapid Prototype interface is too complicated. Multiple windows and query forms pop up on the desktop and subsequently disappear. A tutorial has been designed and distributed with the CD-ROM version to introduce the look-and-feel of the Windows interface, and a UNIX version has been tested on a very localized user group, with good results. Users express satisfaction with the ability to query on spatial, temporal, or thematic criteria, although they criticize the abbreviations used for metadata, which many users find "cryptic". Many users comment about system delays, which will continue to challenge Library use on the Web, where network delays are often unavoidable. The speed of data delivery will continue to challenge system designers. Finally, users acknowledge the need for user evaluation, and their requests that the evaluation mechanism be streamlined will be accommodated in the Web testbed.

SUMMARY AND PROSPECTS

As stated above, the Alexandria Digital Library is expected to become publicly available on the Internet by late Spring, 1996. Issues that must be resolved relate to all four system components. In terms of storage, system designers work to improve the ease of locating archived items, and explorations with wavelet decomposition have created a path by which to reduce the cost of examining maps and imagery in coarse resolution prior to actual downloading. Catalog issues are very important at present, since much of the browsing capabilities are tied to the success of catalog functions. Most placename gazetteers are in non-digital form, and this creates many access problems when dealing with historical maps and references. Another problem relates to merging gazetteers containing spherical coordinates that may be based upon differing Prime Meridians. Many gazetteer entries are associated with somewhat arbitrary point locations, and this challenges automatic generation of spatial footprints. The biggest problem related to ingest issues is the automation of metadata descriptions, which are labor intensive and error-prone when generated manually. These form one of the most expensive aspects of building the Library.

Interface issues focus upon continued iteration between user evaluation and interface refinement to provide easy access to data browsing capabilities for a user population with heterogeneous information needs, if not experience levels. Current efforts in interface design include provision of user-configurable defaults and options, and improving the efficiency of item retrieval and downloading. Planned activities for the user evaluation team will continue to monitor the Web testbed through interactive logging and user surveys. The notepad and associated buttons will be embedded in the Web testbed as well. Plans are underway to solicit hotlists from users to begin generating a communal "hotlist" of pointers to distributed datasets. The biggest challenge for the user evaluation team remains to minimize the gap between evaluation and interface refinement, since it is most efficient to change the interface while it is under design.

Where to find Alexandria

* by telnet (on UNIX with XWindows, this is the Rapid Prototype version)
xhost +
telnet sierra.geog.buffalo.edu
* on the Web (coming soon, will be announced on the homepage)
http//: alexandria.sdc.ucsb.edu/
* if you can't access UNIX or the Web, please leave a business card with
one of the authors at the conference
* email: alexlib@ubvms.cc.buffalo.edu (until June, 1996)

ACKNOWLEDGMENT

This paper forms a portion of the Alexandria Digital Library Project, funded by the National Science Foundation (Grant # IRI-9411330). Funding by NSF is gratefully acknowledged.

Barbara P. Buttenfield and Mark P. Kumler
Department of Geography
Campus Box 260
University of Colorado
Boulder, Colorado 80309
Telephone: (303) 492-3618
email: babs@whitney.colorado.edu

James Barringer, Julian Cone, Robert Gibb, Hamish Heke, David Medyckyj-Scott, Peter Newsome, and Janice Willoughby

Environmental Data Access in New Zealand 1985 - 1995: user-pays to open-access and the WWW

ABSTRACT

In 1986, political reforms moved New Zealand into a period of "user-pays" economics. This has had significant effects on data ownership, data access and data cost for Government-held environmental data, both spatial and non-spatial. Policies encouraging "cost recovery" by Government research organisations led initially to heavy data charging, but this was found to be untenable. More recently a more open "cost-of-supply" policy has become established, and data suppliers are now ready to take advantage of the rapid developments in the INTERNET and WORLD WIDE WEB (WWW) environment. Such developments, although largely welcomed by both data providers and users, have raised new issues centred around networked database access, data interpretation and presentation standards, and links to other databases.

Landcare Research maintains New Zealand's primary land resource database, the New Zealand Land Resource Inventory. Until recently, use of the database has been hampered by data charges, limited GIS staff and facilities, and difficulty for non-expert users in interpreting the database. To resolve these issues Landcare Research has developed a WWW user interface which allows users with no GIS skills to query the spatial database interactively, and see the results on screen in GIF format, or in hardcopy. The interface structure is designed for maximum flexibility, to cope with internal database changes, as well as new links to external databases. Analyses and outputs are standardized to limit inappropriate uses of data by users.

BACKGROUND

In response to an economic crisis, major and complex political and economic reforms beginning in 1986 had far reaching effects in all elements of New Zealand's political, economic and social structure. In Government-run research organisations three key changes occurred. First, a drive to reduce Government spending led to both reductions in funding and the introduction of "users pays" policies which required that research organisations recover a proportion of their costs through consultancy and sales of services or goods.

Second, the Government restructured science in New Zealand to introduce greater accountability, and to make more efficient use of resources directed at well focussed research. This started in 1986 with a complex series of organisational amalgamations, but later involved major restructurings. Eventually the 20 divisions of the Department of Scientific and Industrial Research (DSIR), and research divisions of four large government departments

(Works & Development, Forestry, Agriculture & Fisheries, and Meteorological Service), were reorganised into nine Crown Research Institutes (CRIs) which now carry out most publicly funded research in New Zealand, other than medical research.

Third, this science restructuring process also involved major changes to the funding system from bulk funding to an output oriented competitive funding system, the Public Good Science Fund (PGSF), administered by the Foundation for Research Science and Technology (FRST). CRIs, universities and research associations all bid competitively for funding through this system.

Of these three changes, the first, in particular, generated financial imperatives which had a significant effect on the perception of spatial and non-spatial databases as major assets. This, in turn, generated debate over data ownership, data access and data cost for Government held environmental databases.

This paper has two parts. First, it discusses the impact that political changes have had on collection of, and access to, environmental data over the last decade, and flow-on effects on GIS-based environmental modelling in New Zealand. Second, it looks at the future for data access in New Zealand, and also at Landcare Research's Geographic User Interface to Landcare Databases (GUILD-on-the-Web) which is leading the way to more open and accessible environmental databases in New Zealand.

DATA ACCESS IN THE 1980s

In 1985, GIS was not widely used in New Zealand. The Ministry of Works and Development, Water and Soil Division (MWD/WS) had started in 1977 with its own series of GIS for storing, analysing and plotting the New Zealand Land Resource Inventory (Williams, 1985; Van Berkel & Williams, 1986), but would move to ARC/INFO in 1988. The Department of Lands and Survey had also been using CAD and GIS since 1977, first to support the development of the metric version of their cadastral database (DCDB), and later for a new metric topographic map series (DTDB). Very little other digital data, particularly spatial data, was available at that time, however much data was available in hardcopy (eg. DSIR soil and geological maps), or text based digital databases (eg. New Zealand Meteorological Service Climate Database). In this "traditional" (pre-GIS) setting, expensive data collection was well supported by public funding, there was nominal recovery of publication costs through map sales, and "data" was freely available to the public and user community with few practical limitations (Giltrap et al., 1993).

In 1986 the "user-pays" principal was applied to research organisations. Virtually overnight, data, whether spatial or non-spatial, digital or hardcopy, became a valuable asset for research organisations which were suddenly required to recover a fixed proportion of their appropriation from commercial revenues (Newell, 1992). The nature of this newly created "data market" was a series of virtual monopolies, each type of data being available from, and controlled by, one source (eg. DSIR Land Resources for soils and land resource data, and the Department of Survey and Land Information for topographic data). An immediate effect was to restrict data use, particularly for digital data, through the introduction of substantial charges for data. This created a dilemma for the data suppliers. While there was a requirement to generate income from these data assets, and considerable effort was being expended on developing digital databases to make the data more readily available, data suppliers tended to price themselves out of the market, despite the fact that in most cases there was no other alternative source from which to obtain similar data. Many data users simply could not afford the prices being asked, and did without.

So, at a time when computer technology was developing rapidly, enabling development in GIS functionality on increasingly cheaper and more powerful hardware platforms, availability of spatial data was severely restricted, primarily by cost. This and other factors combined to restrict users ability to integrate useful datasets for modelling spatial processes. GIS-based environmental modelling activities with complex data requirements were rare, and simple interpretations based on classifications of single data layers were more common.

This situation persisted throughout the late-1980s. By 1990 it had become clear that heavy charging for data was untenable, and data providers found that their data was being under-utilised. A Ministerial Science Task Group, set up to advise on the restructuring of science provided some recommendations for managing databases: the Crown should fund these databases, retain copyright to the data, but make them publicly available on the basis that the spatial information had been collected using public funds. However, data charges should remain where the costs of collection, archiving and maintenance were not covered from public good funding. The costs of actual retrieval of information from databases and collections could also be recovered (MSTG, 1991). These recommendations moved the principles of data access and charging more into line with those that had applied in the pre-GIS era of the traditional sale of hardcopy maps.

DATA ACCESS IN THE 1990s

Since the formation of the CRIs in 1992, data access has steadily improved. The PGSF supports collection of data having substantial scientific merit, and has identified "nationally significant databases", for which base funding levels have been identified, to ensure that key databases are at least maintained (FRST, 1993). Current funding levels remain above this base level, but funding for some database updating can still be difficult to obtain. However, the recovering New Zealand economy, along with the role defined for Regional and District Councils by the Resource Management Act, and the removal of direct data charges, is generating an increasing demand for data from the Regional and District Councils. Collaboration between organisations to jointly fund national databases is also beginning to occur (eg. a consortium is negotiating to develop a land cover database from satellite imagery).

The PGSF focus on scientific relevance has also resulted in a move away from predominantly single layer vector format databases with classified attribute data collected by field survey. For a number of years, the limitations for modelling and problems of updating these databases (eg. NZLRI) have been well recognised, but the step to a more multi-layered single-factor database model has been difficult to initiate. However, more focused funding, scientific requirements for objectivity and mechanistic models, and the increasingly multi-disciplinary nature of much of the research being undertaken, has encouraged the move towards single factor raster format databases and greater reliance on modelling, both to derive layers from DEMs or remote sensing and to develop more defensible analyses.

These processes have now returned most data suppliers to a point where data access is limited mostly by the physical constraints of data provision. The key problem currently being addressed is how to meet the growing demand for data with the resources available. Landcare Research is no exception, and has had a project running for three years to develop tools to meet this need.

GUILD-ON-THE-WEB

Geographic User Interface to Landcare Databases (GUILD) is a natural development from an earlier project for a geographic index of databases held by Landcare Research (GILD). In 1992, Landcare Research was formed from parts of seven previous organisations, all of which held their own spatial and non-spatial databases. Knowledge of what databases were available, and what information they held, was often poorer within Landcare Research's internal research community than for external data users who had a history of contact with one or more of the preceding organisations. Given recent increases in demand for information to solve environmental, economic, and social problems, the aim of GILD was to provide Landcare Research staff with descriptions of the data available within the organisation. GILD was implemented in ARCVIEW v1 and displayed the spatial distribution of each database record of the major databases held by Landcare Research, together with their associated meta-data on age, content, resolution, quality and accessibility. For databases associated with physical specimen collections such as herbaria and insects, the level of spatial detail in the metadata meant that the collection location of each specimen was displayed.

As database access was freed up in New Zealand during 1992 and 1993, it became clear that, while GILD had considerable merit in its own right, there was significant scope for a more ambitious project. This was also driven by the recognition that Landcare Research's limited resource of trained GIS staff would not cope with a significantly higher level of research-related database query tasks in the newly restructured company, and that a mechanism giving staff direct access to the GIS was needed. The high proportion of requests related to research applications that were being treated as one-off requests rather than standard data dumps or queries did nothing to reduce this bottleneck in staff resources. The GIS group determined to "grow" GILD from a simple geographic index to a full GIS user-interface that could deliver solutions directly to users who had varying familiarity with the data and often negligible knowledge of GIS technology. Initial plans to use ArcView 2 and AVENUE as the software platform for this interface were changed because of the hardware requirements of ARCVIEW 2, which on release were substantially greater than the staff standard of a Novell networked 486DX66-2 with 8MB of memory.

Three options were examined:

- X11 emulation on MS-Windows to a remote GIS session.
- Genamap's Genius 2 Windows GUI converter, which can support an X11 GUI interface to a GIS which can be ported to a MS-Windows environment as a single executable that runs without the need for a full X11 emulation.
- The World Wide Web (WWW).

The first option suffered from constraints similar to ArcView 2, in that it requires more hardware resources than are typically available, and a learning curve inappropriate for people who otherwise have no need for UNIX access, and only occasional use for the system. The second option looked very promising but at the time (late 1994) had not been released. The third option appealed because of the potential to open access to an even wider audience and allow them to use very ordinary PCs with readily available software already in wide use for other tasks. The biggest drawback was the restricted set of interactive widgets supported on the WWW by HTMLv2. An analysis of the user interface requirements and careful initial design indicated, however, that a solution was feasible using the HTMLv2 plus some additions from HTMLv3.

A strategy was therefore adopted to go with a WWW user interface supported by Genamap scripts which would be developed so that they could also be called from a Genius 2 user interface. This would leave a future option of using WWW for basic users and for supporting users with more complex requirements, either through WWW down loadable applets such as Java or via a down loadable GENIUS 2 executable under MS-Windows. More recent developments have confirmed the strategy as being the right one with Java moving rapidly ahead.

In designing the system, we have used the following fundamental philosophies:

- to hide as much of the GIS technology as possible
- to present users with an information-oriented user-interface
- to foster the concept of GIS standards in the generation of all outputs
- to encourage appropriate use of data through keeping the user informed of inherent data limitations.
- to constrain users so they are only presented with reasonable options.

The approach used to achieve this has been to remove all data dependancy from the code and place it in a single actionable metadata store.

The new interface called GUILD-on-the-Web, concentrates initially on delivery of relatively simple GIS processes from a wide range of data sources held by Landcare Research. At present a beta version is available to all Landcare Research staff for approval and testing. It is planned to promote it to progressively wider audiences over the next 12 months.

STRUCTURE OF GUILD

GUILD-on-the-Web comprises three primary modules: a WWW front end based on cgi scripts, a metadata and metaprocess system, and a GIS application back end. The basic process involved in using GUILD is to make a series of selections within the WWW front end query forms. The options made available to each user are acquired by querying the metaprocessing system. The final selections are coded and sent to the host GIS, at present Genamap, which queries the metaprocess system to find out how to implement each component of the total request, and then performs the requested transactions. At present, the system caters only for graphical output. Once the GIS has completed the task defined by the metaprocessing commands the results are returned to the remote site for viewing by the user. CLICK HERE, and you can study the GUILD-on-the-Web home page. Access to help pages, future plans and registrations of interest is enabled, but full access to GUILD is currently only available to Landcare Research staff.

This paper does not describe the technical details of the metaprocessing system which will be the subject of another paper. However, the key point of this overall structure is to achieve maximum flexibility. The metaprocessing system is the key to this flexibility. Stored in this system is a database defining everything, from the layers available for query, to the GIS functions that may be used. The aim is to ensure that the WWW front end is generated

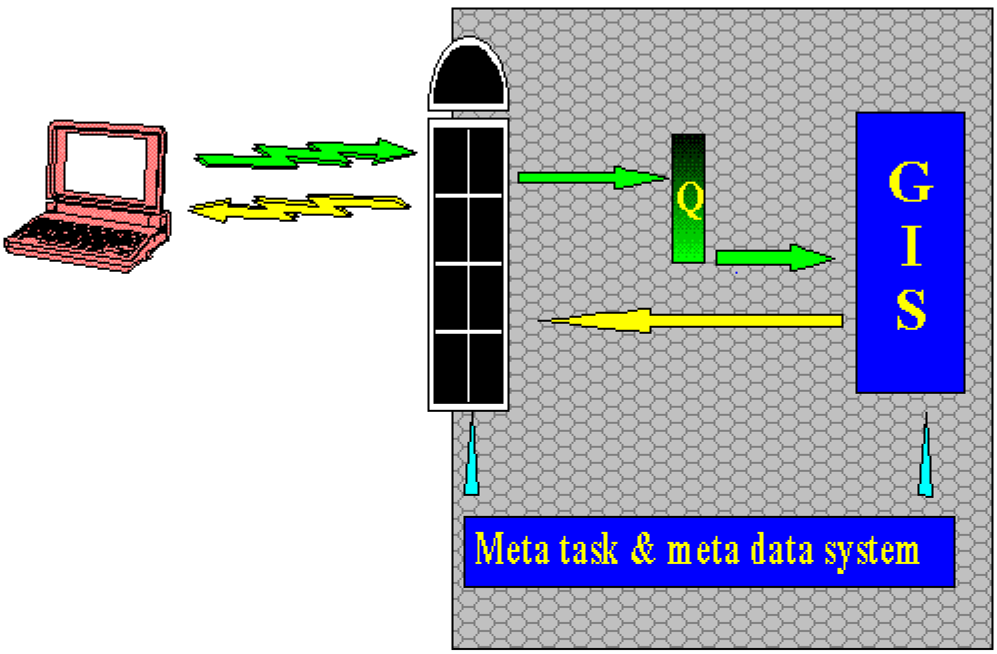


Figure 1: simple schematic diagram of GUILD setup to a GIF format and returned, via the metaprocessing system to the WWW front end and the remote user.

"on-the-fly" from this system, so that new data layer additions and analytical capabilities can be added to the overall system without the need to recode the front end. One further advantage of the metaprocessing system is that the data it contains is structured to be independent of any particular GIS. In addition to Genamap, an ARC/INFO back-end implementation is also under development.

THE FUTURE OF GUILD

Three development paths will be followed in parallel over the next twelve months: Continued assessment and incremental improvement of the user interface; provision of additional functionality of both the GIS back end and the user interface; and extension of the range of data sources and interpretations that are delivered through GUILD-on-the-Web.

The key elements of the functionality improvements are delivery of tabular reports, point and click map query, keyword data exploration capability to identify related data held by Landcare Research, hard copy map ordering and delivery service, support for raster data sources such as satellite imagery, aerial photography and digital elevation models, and access to environmental models. These and other improvements are described in the Futures page accessible from GUILD-on-the-Web.

CONCLUSION

The results of new economic policies and the restructuring of science in New Zealand have had a major impact on access to both digital and hardcopy sources of environmental, topographic and other data over the past decade. Unfortunately, these events occurred during a period of rapid technological development in computers and GIS, which was also a period of rapidly developing need for access to the full range of data types necessary for resolving complex multi-disciplinary research problems. The result has been that development of GIS as a tool for environmental modelling has been restricted in New Zealand over the last ten years.

New Zealand's economic recovery in the 1990s, along with recent changes in data charging policies, Government PGSF funding priorities, and the development of the WWW have provided an environment which encourages more open access to data, increased collaboration between data suppliers and data users (both in research and land management), and development of more flexible database structures. In this more open environment, Landcare Research has developed its GUILD-on-the-Web user-interface to provide access to all of it's spatial databases. The growing demand for data access has already resulted in plans to extend the GUILD concept to the digital databases of three other major New Zealand environmental data suppliers.

ACKNOWLEDGEMENTS

Thanks to Grant Hunter, Anne Austin and Ian Whitehouse for editorial comments on drafts of this paper. Development of GUILD is funded by the Foundation for Research Science and Technology under contract C09518.

REFERENCES

Foundation for Research Science & Technology (1993). Nationally Significant Public Good Science Fund Databases and Collections, Foundation for Research Science & Technology, Wellington, New Zealand, September 1993, 22pp.

Giltrap, D.J., Gibb, R.G. & Newsome, P.F. (1993). Natural Resource Databases - Progress and Trends, Fifth Annual Colloquium of the Spatial Information Research Centre, 17th - 19th May, 1993, University of Otago, Dunedin, New Zealand, p43-55.

Ministerial Science Task Group (1991). Crown Research Institutes: research companies for New Zealand - the report of the Ministerial Science Task Group. 227pp.

Newell, J.O. (1992). Operation of Spatial and Statistical Information Markets in NZ: roles of the public and private sectors, Fourth Annual Colloquium of the Spatial Information Research Centre, 18th - 20th May, 1992, University of Otago, Dunedin, New Zealand, p25-46.

Williams, R.D. (1985). LADEDA Beginners Handbook, unpublished user manual, Ministry of Works & Development, Christchurch, New Zealand, 155pp.

Van Berkel, P.R. & Williams, R.D. (1986). TRACE/POLAR Users Manual, Publication No. 8 of the Hydrology Centre, Ministry of Works and Development, Christchurch, New Zealand.

AUTHOR INFORMATION

James Barringer, Julian Cone
Landcare Research New Zealand Ltd.
P.O. Box 69
Lincoln
NEW ZEALAND
Phone: +64 (3) 325 6700
Fax: +64 (3) 325 2418
email: barringerj@landcare.cri.nz

Robert Gibb, Hamish Heke, David Medyckyj-Scott, Peter Newsome & Janice Willoughby
Landcare Research New Zealand Ltd.
Private Bag 11-052
Palmerston North
NEW ZEALAND
phone: +64 (6) 356 7154
fax: +64 (6) 355 9230
email: gibbr@landcare.cri.nz

Kate Beard

A Structure for Organizing Metadata Collection

Abstract

Interest in sharing data, providing network access, and in documenting data quality has fostered much of the recent interest in metadata and development of metadata standards. Metadata should provide a full description of data such that it can be discover ed by potential users, assessed for its usefulness, transferred, and used or analyzed in an appropriate context. In the environment of electronic data sharing metadata is essential and it is important that it be as accurate as possible. The compilation of metadata requires as much if not more care as compilation of the original data.

Metadata as it is currently being compiled has two limitations: 1) it is largely being compiled manually after the data has been collected, and 2) it has been treated as separate from the data The former problem is expected since the concept of metadata has only recently emerged. The consequence is that retroactive compilation of metadata is an error prone and sometimes even impossible task. With respect to the latter problem, typical formats for metadata are separate text files, or separate relational tables. This paper proposes a structure for organizing metadata collection that moves it toward greater automation and integration with the data it describes. This work is described in relation to metadata for marine data and associated models.

Introduction

Metadata fundamentally describes data. In this paper it is defined as a formal description of data that allows them to be exchanged and used by people other than those who originally collected them. This paper also addresses metadata for models. In the co ntext of digital libraries metadata for both data and models becomes essential. The questions which arise are where does this all important metadata come from and how is it most effectively organized. To develop a formal structure for metadata, this paper considers the following questions: 1) what constitutes metadata, 2) who is responsible for generating metadata, 3) when is metadata most optimally generated and 4) how can it be generated most efficiently?

What Constitutes Metadata?

The question of what constitutes metadata has been the subject of much debate and effort. Some content lists are long and others short. The Content Standard for Geospatial Metadata (FGDC 1994) represents one outcome of efforts focused specifically on spa tial data. The Dublin Core (Weibel, et al 1995) developed by the library community represents another. Bretherton (1994) as part of an IEEE workshop provides an interesting white paper on a metadata reference model. Smith (1995) describes metadata as cons isting of context, content, and structural information. All of these metadata efforts show substantial agreement in terms of basic metadata content.

The question of what constitutes metadata is re-structured here on the basis of four functional roles: 1) search, 2) retrieval, 3)transfer, and 4) evaluation. These functions provide a useful structure to organize metadata content. They are not intended as exhaustive or mutually exclusive classes. In particular this structure can help to isolate the metadata which forms an independent catalogue entry and what metadata must be integrally linked with the data and travel with in when it is retrieved.

With respect to the search function, metadata should provide sufficient information to discover if the data of interest exists within the collection of available data or exists at all. With respect to the retrieval function, metadata should provide the information for users to acquire the information of interest. The library analogy for this is the procedure for checking out a book. The retrieval component of metadata may be as simple as providing a URL identifying the location of an electronic data se t, to as complex as covering security issues or arranging a financial transaction for access to the information. It may include such information as the off-line location of the data, the contact person, media formats for data distribution, any restriction s on access to the data such as licensing agreements and information on costs. Metadata to support transfer should provide the necessary information for users to make use of retrieved files on their machines. This component would include information on the size of the data set (and its metadata), and the logical and physical structure of the data and metadata. The evaluation function of metadata is perhaps the most complex. Metadata to support evaluation can consist of any information which assists u sers in determining if data will be useful for an application. In many ways it is a refinement or expansion of the other functions. Some evaluation should be possible prior to retrieval particularly if there are fees for the data. In addition sufficient metadata should travel with retrieved data such that evaluation can occur or continue subsequent to retrieval. The next sections focus on the search and evaluation functions as these are the most pertinent to scientific data analysis and modeling.

Metadata for Search

Metadata to support the search function includes any information which would help a user discover an information resource. We think of author, title, and subject indices as the traditional metadata elements for a library search, but for spatial data thes e are typically not the optimal search indices. To set the scene we should stipulate the type of archive we plan to search. For the purposes of this paper a digital library of scientific data collections and models is assumed. These collections could inc lude imagery, maps, in situ collections, models, and related bibliographic materials in a heterogeneous distributed archive maintained by several investigators. Heterogeneous is understood to describe data sets which vary in their quality, area of spatial coverage, range of temporal coverage, their spatial and temporal resolution, their generating source, database schema, and other characteristics. Given this scenario we want to determine the metadata required to effectively search such an archive.

The search indices to such a collection should be multidimensional and include various spatial, temporal, and thematic indices. A spatial index can take several different forms which might include an index for two dimensional space, three dimensional sp ace, topological relations and metric relations. The temporal index might include calendar and clock times as well as process time (tides) and temporal relationships. Thematic indexes should allow searches on any type of data collection, data collector, t hematic variable as well as measures of variable similarity. Table 1 provides a potential list of metadata elements for search.

Table 1. Summary of Metadata Elements for the Search Function

Spatial Indices

- Geographic region
- Vertical range
- Horizontal position
- Vertical position
- Topological, metric relationships

Temporal Indices

- Collection period
- Calendar/clock time
- Process time
- Event relationships

Thematic Indices

- Collection name
- Data type
- Variable name
- Related variables
- Keywords
- Variable description
- Data collectors/authors

Search should theoretically be possible on any metadata element so all metadata may be considered relevant to search. Strictly search metadata includes the information required to find a data set which meets a set of criteria, but it can overlap with eval uation. In some senses a data set is evaluated by its ability to satisfy certain criteria so in meeting a set of search criteria it may have met a minimum set of evaluation criteria as well.

Metadata for Evaluation

If users wish to evaluate the utility of data for a particular application we can ask what is the pertinent information which should be supplied to them. Important information likely includes the purpose for which the information was originally collecte d, the method by which it was collected, and structural information describing the data.

Given the question: Is the data likely to be of use to me? Bretherton (1994) suggests metadata should include the following pieces of information.
(2) A summary description of scientific context, including discussion of primary scientific objectives for the data, the variables, instrument systems, processing algorithms and quality control procedures
(3) The spatial, temporal coverage and sampling design
(4) The scientific credentials of this data, including evidence for its credibility, references to scientific publications which used it or commented on its quality or deficiencies.

For the question: Is it really what I want? Bretherton (1994) suggests metadata in the form of the following browse products: (1) a typical sample, (2) diagrams, (3) graphs, and (4) derived products. Bretherton (1994) envisaged summaries of this informat ion, prepared as electronic documents with linkages to various layers of information containing more detail if desired.

Other examples of metadata for evaluation are provides by the US Global Change Research Program (USGCRP), US GLOBEC and related programs. The guiding principle in their data management policy is, "as soon as data might be useful to other researchers the d ata should be released along with documentation which can be used by other researchers to judge data quality and potential usefulness."

The USGCRP data policy requires that all principal investigators submit 1) pre-data collection plans that describe in detail collection and analysis methodologies, 2) a detailed inventory of all measurements actually made along with documentation of the m easurement techniques used to produce the data, 3) an estimate of the accuracy and precision of each measurement along with procedures used to correct errors, remove noise, or otherwise modify the collected data and any analyses of the data 4) documentati on of the physical setting of the ecosystem and 5) corrections and improvements in data made subsequent to submission of the data to the data management office. Table 2 lists a synthesis of metadata elements for evaluation.

Table 2 Metadata Elements for the Evaluation Function

Process

- Sampling plan
- Instrument name
- Instrument variable

Instrument operation
- Quality-reliability
- Processing history

Context

- Physical

Browse Files

- Sample plan diagram
- Images
- Maps
- Graphs
- Derived products

Use

- who used
- for what purpose
- citations

Who Should be Responsible for Generating Metadata

Historically librarians played the dominant role in generating metadata. They performed the valuable function of abstracting, indexing, and cross-referencing information such that it could be efficiently discovered. In the case of scientific data exchang es, data collectors or principal investigators were more often the primary generators of metadata. Early (pre digital and even early digital) exchanges were generally infrequent. The two parties to an exchange could negotiate the transaction one on one a nd relay the appropriate information about the data given the context of the exchange. As data exchanges become more frequent and eventually routine they will become too burdensome for data collectors to individually negotiate. Procedures for exchange wi ll need to be formalized such that information can be transferred without a need in every case for person to person interaction. Such exchanges will require sufficient information to serve the four functions discussed above. The possibilities for metada ta generators or compilers include some of the same historical players; the librarian and the data collector or domain expert but may expand to include new players such as computer scientists, electrical engineers, or others involved in knowledge extracti on and data mining. Librarians have an important role as the traditional experts in abstracting and cataloguing information. The data collector and domain specialist have the familiarity with the process of data collection and expertise in the characteri stics of a particular type of data. Within this process there will likely be complementary or collaborative roles for each type of player. The non-human extensions to this include smart data loggers and linked instrumentation as well as smarter systems t hat are capable of logging metadata as data are processed.

When Should Metadata be Generated or Compiled?

Metadata need not be collected all at once and there are in fact distinct points in the history of a data set at which metadata is logically generated. Times at which metadata collection may occur can be identified as three broad stages: pre-data collect ion; collection concurrently with data collection; and post-data collection. Characteristics of the data and metadata will determine the optimal stage. This three stage process of data collection is now being recommended by the USGCRP. The individual stag es are discussed in more detail below.

Pre-data collection. Pre-data collection of metadata can occur in two forms: that which is not specific to the data to be collected and that which is. Non-specific metadata compiled prior to data collection could include such things as thesaurus, gazettee rs, and instrument descriptions. Metadata specific to a particular data set which could logically be compiled prior to data collection includes the data sampling design methodology, specification of purpose, planned geographic coverage, planned depth or elevation ranges, planned collection period, and instrumentation. As an example from marine data collections, the US GLOBEC pre-collection reports require inclusion of: study objectives, principle investigators, sampling plan, identification of data type s and instrumentation. Typical descriptions include navigation, timekeeping, sensor make and model, net opening, mesh size, rate of retrieval, mooring configuration and other information particular to a data collection device.

Concurrent data collection. Concurrent data collection occurs simultaneously with data collection. Concurrent metadata would be supported by linked instrumentation in which at the same time a sample is recorded other variables are recorded such as horiz ontal position, vertical position or depth, time, salinity, temperature, wind speed, and instrument variable settings The U.S GLOBEC Steering Committee strongly recommends the use of logging systems which record the underway data including navigation, met eorology, near surface temperature and salinity, and any other data collected automatically.

Post data collection: Post data collection of metadata includes compilation which could only logically take place after data has been collected. These elements would include the actual processing history, history on use of the data, quality assessment, g eneration of browse files, and computation of additional indexing, such as computation of topological relationships, and identification and indexing of image content.

Methods for Metadata Generation and Compilation

There are several options for metadata compilation. In the ideal case we would wish to automate as much of the metadata collection and compilation as possible. The methods will in most cases be dependent on the time of compilation (pre, concurrent, or po st) and on the data collection methods. The range of possible compilation methods include 1) key in, 2)look-up, 3) measured, 4) computed and 5) inferred. Key-in is the least desirable since it is tedious and error prone. For some metadata elements, howe ver, this will be the only option. The look-up option assumes information has been previously entered and a new element is determined by using a related element value to look up an appropriate code. For example, assuming a named geographic region (the Gul f of Maine) has been keyed in, the appropriate bounding coordinate box for this region can be found in a look up table or gazetteer. This could also work in reverse, the measured coordinate bounding box could provide the key to look up a named geographi c region. Actual measurement of metadata elements should provide the easiest and potentially most error free method. Examples of measured metadata elements would include electronic measurements of horizontal or vertical position using GPS, or measuremen t of physical context information such as temperature, wind speed, or other pertinent variables at the same time the variable of interest (e.g. water quality) is observed. Concurrent collection required linked instrumentation or "smart" data collectors. A large volume of metadata may also be computed from other metadata elements or the data themselves. Example of computed metadata elements could include horizontal positions computed from navigation instruments and time, vertical positions computed from stereo aerial photography, topological or metric relationships computed from coordinate data, or some quality measures computed from observed data. Inferred metadata elements would include those inferred from one or more other metadata elements and/or the data. Examples of inferred metadata may include an inferred time period based on characteristics of the data such as temperature, color, or behavior. Inference may in many cases be the only method possible for retroactive compilation of metadata. Infer ence of metadata elements overlaps to some degree with data mining or knowledge discovery (Frawley et al 1991) which is the extraction of implicit, previously unknown information from a database or collection of databases.

Metadata for Models

Models and models outputs that are shared electronically will also require metadata. The metadata for models can have a very similar structure to metadata for data. The functional roles are similar. We still need to search for, retrieve, transfer and eval uate models and model outputs. The stages for compilation will be similar to those suggested above for data but will translate to 1) pre-model run compilation, compilation concurrent with a model run and post-model run compilation. The compilers of meta data for models will largely be the model authors but could also include domain specialists. The compilers of the metadata for model output will logically be the models themselves and this compilation would logically occur concurrently with a model run. The primary compilation methods will be key in of information by authors or computation of metadata by the model itself.

Formulating a Structure for Metadata Collection

In this section information on the metadata elements, compilation stage, compilers and compilation methods are combined to outline a structure and strategy for metadata collection. In each of the following tables the metadata elements are listed along wit h the compilation stage, primary compiler and compilation method. Table 3 illustrates a structure for metadata elements for search.

Table 3. Organization of Metadata for Search

Metadata element	Compilation stage	Compiler	Compilation method
Spatial			
Geographic region	Pre/Post	Data col, Lib/Sy	Key, lookup/comp, infer
Depth range	Pre/Post	Data coll/Sys	Key in/compute
Horizontal position	Concurrent	Instrument	Measure
Vertical position	Concurrent	Instrument	Measure
Topological relationships	Post	System	Compute
Metric relationships	Post	System	Compute
Temporal			
Collection period	Pre/Post	Data col/system	Key in/compute, infer
Calendar, clock time	Concurrent	Instrument	Measure
Process time	Pre/Post	Domain Spec	Key in/compute, infer
Event relationships	Post	System	Compute, infer
Thematic			
Collection name	Pre	Data Co, Dom	Key in
Data type	Pre	Data Co, Dom	Key in
Variable name	Pre	Data Co,Dom	Key in
Related names	Pre/Post	Domain Sp/Lib	Key in/lookup
Keywords	Pre/Post	Domain Sp/Lib	Key in/lookup
Variable description	Pre	Data Co, Dom	Key in/lookup
Data collector, author	Pre/Post	Data collector	Key in

It is not efficient to describe each element in the table but a few examples will illustrate the intent of the structure. The back slash which separates values in the table is used to indicate that the terms following the back slash are related. For examp le in the case of the first element the back slash indicates that the compilation methods, key in and look up, apply to the pre compilation stage and the compilation methods, computed and inferred, apply to the post compilation stage. Referring to the fir st metadata element "Geographic region", the possible compilation stages are pre or post collection. Most data collection campaigns will have defined the general geographic region prior to data collection so this element could be keyed in by the data coll ector. To expedite metadata collection of similar pre-data collection elements, interactive pre-collection forms could be designed for easy key in by the data collector. Geographic region could take two forms, a named geographic region or specification b y coordinates. If either one has been keyed in the other can be generated through a look up table or gazetteer. Librarians are included in the list of compilers for their potential role in generating a look up gazetteer. In the case of post data compila tion a coordinate defined geographic region may be computed from the set of measured coordinates. For example if a cruise visits fifty stations and a position is measured for each station, the bounding rectangle or convex hull can be computed for this poi nt set to generate a value for the geographic region. In this case the compiler is the system. If this metadata element is to be compiled retroactively and no geographic measurements exist for a data set it may still be possible to infer a geographic re gion from characteristics of the data set. MacGranaghan) The compiler could be a system, but a highly specialized knowledge extraction mechanism would be required in this case. The depth range and the temporal collection period metadata elements have sim ilar structures.

As additional examples, the horizontal and vertical position elements are indicated as being collected concurrently with the data. The logical approach would be through linked instruments in which for example horizontal and vertical GPS coordinates are g enerated at the same time a variable is observed. The compiler is thus an instrument and the metadata element is directly measured. The calendar/clock time element has similar behavior. Given these examples the structure of the other elements in the tab le should be self explanatory.

Table 4 illustrates the same structure for metadata elements for the evaluation function. The sampling plan element is one which should be documented prior to data collection. It is most logically compiled by the data collector and by key in on some type of interactive pre-compilation form. If the sampling plan has a spatial configuration a sample plan diagram can be included as a browse file. This element would also be compiled prior to data collection by the data collector or domain specialist. Key i n may be the only current method for compiling this element but the possibility exists for computing a spatial sampling plan as well. Most of the other browse files would logically be compiled after data collection. However there is the possibility tha t some could be compiled concurrently with data collection. Examples would be underwater video images or ship tracks which could be plotted as the ship steams from station to station in the process of collecting data.

Table 4. Metadata Elements for Evaluation

Metadata for Evaluation

Metadata element	Compilation stage	Compiler	Compilation method
Process			
Sampling plan	Pre	Data co,Dom	Key in
Support	Pre	Data collector	Key in/lookup
Instrument name	Pre	Data collector	Key in/lookup
Instrument variable	Pre/Concurrent	Data collector	Key in/lookup
Instrument operation	Pre/Concurrent	Data co/Inst.	Key in/measured
Quality-reliability	Concurrent/Post	Data co,Dom	Key in/compute,infer
Processing history	Post	Data co/Sys	Key in/compute
Context			
Physical	Concurrent/Post	Data co/Sys	Measure/infer
Browse files			
Sample plan diagram	Pre	Data co, Dom Sp	Key in,compute
Images	Concurrent/Post	Instrument/Sys	Measure/compute
Maps	Concurrent/Post	Instrument/Sys	Compute
Graphs	Post	System	Compute
Derived Products	Post	System	Compute
Use			
Who used	Post	User/Sys	Key in/measure
What Purpose	Post	User/Sys	Key in
Citations	Post	Librarian	Key in

Tables 5 and 6 present the metadata structure for models and models outputs. Elements which are independent of the model run would include such items as the name, type of model and type of process being modeled, parameters, boundary conditions, authors, citations of the model, and model software. Information which could be recorde d concurrently with a model run could include actual parameter values, boundary condition values, spatial units and temporal units, and pointers to the observational data used to force the model. If the intent is to evaluate a model outcome, it is likely that an investigator may wish to evaluate the data used to force the model. Thus a useful component of model metadata will be a link to the metadata of the observational data used to force the model. Post model run metadata could include evaluation metho ds and results, references, and browse graphics of the model output.

Table 5. Metadata Elements for Models and Model Results - Search

Metadata element	Compilation stage	Compilor	Compilation method
Model & results			
Name	Independent	Author	Key-in
Type of model	Independent	Author	Key-in
Type of process modeled	Independent	Author	Key-in
Parameters	Independent	Author	Key-in
Authors	Independent	Author	Key-in

Table 6. Metadata Elements for Models and Model Results- Evaluation

Metadata element	Compilation stage	Compilor	Compilation method
Model			
Authors	Independent	Author	Key-in
Publications	Independent	Author	Key-in
Software	Independent	Author	Key-in
Parameters	Independent	Author	Key-in
Boundary conditions	Independent	Author	Key-in
Citations	Independent	Librarian/Sys	Key-in/Compute
Model results			
Parameter values	Concurrent	Model	Key-in
Bounary condition values	Concurrent	Model	Key-in
Forcing data + conditions	Concurrent	Link to data metadata	
Spatial units	Concurrent	Model	Compute
Temporal units	Concurrent	Model	Compute
Validation data	Post-run	Link to data metadata	
Model output images	Post-run	Model	Compute

Summary

The goal in developing this structure to examine and organize metadata was to improve the overall process for collecting and compiling metadata. Good metadata is essential for the extended use of data, models and model results by wider numbers of resear chers and for extending the useful life of the data. The current retroactive and haphazard compilation of metadata does not support these efforts. The structure developed in this paper should help to organize our metadata collection strategies by thinkin g about optimal staging for metadata compilation and introduction of more tools to automate the process. The tables indicate key in of metadata elements in many places which is reflective or our current capabilities. Ideally we would like to minimize the number of key in entries. Outlining metadata in this structure helps provide a focus for directing automation efforts.

Another interesting side benefit of this structure is that it points to a need for metadata about the metadata. For example in an evaluation context it may be important for a user to know the compilation method of a metadata element. For example if a geog raphic region or time period was inferred for a data set this provides important information about the reliability of that element. This complexity quickly indicates that traditional relational database models will not be adequate for metadata representa tion.

References

Bretherton, F. 1994. Reference Model for Metadata : a Strawman. http://www.llnl.gov/liv_comp/metadata/papers/whitepaper/bretherton.ps

FGDC 1994. Content Standards for Digital Geospatial Metadata. June 8. Federal Geographic Data Committee. Washington DC.

Frawley, W. J., Piatetsky-Shapiro, G., Matheus, J. 1991. Knowledge Discovery in Databses: An Overview. in Knowledge Discovery in Databases, AAAI Press Menlo Park. 1-27.

Futch, S., Chin,D., McGranaghan, M., and J-G. Lay. 1992. Spatio-Linguistic Reasoning in LEI (Locality and Elevation Interpreter). in A. Frank, I. Campari, and U. Formentini (Ed). Theories and Models of Spatio-Temporal Reasoning in Geographic Space. Pisa , Italy 318-327.

Smith, T. R. 1995. Paper presented at DL Metadata Workshop. Santa Barbara.

U.S. GLOBEC. 1994. Ocean Ecosystems Dynamics. U.S GLOBEC Data Policy Report Number 10. February 1994. http:www.ccpo.odu.edu/globec/globec_rn_10_feb_1994.html

Weibel, S., Godby, J., Miller, E. and R. Daniel. 1995. OCLC/NCSA Metadata Workshop Report. Office of Research Online Computer Library Center, Inc. http//www

---

Kate Beard

NCGIA

Department of Spatial Information Science and Engineering

University of Maine

Orono, ME 04469

Beard@spatial.maine.edu

Next

Next: Introduction

UDK: A European Environmental Data Catalogue

Oliver Günther
Institut für Wirtschaftsinformatik
Humboldt-Universität zu Berlin
Spandauer Str. 1
10178 Berlin, Germany
guenther@wiwi.hu-berlin.de

Helmut Lessing and Walter Swoboda
Niedersächsisches Umweltministerium
Archivstr. 2
30169 Hannover, Germany

Abstract

In this paper we give an overview of the UDK project, an international software engineering effort to facilitate access to environmental data. The UDK (Environmental Data Catalogue) is a metainformation system and navigation tool that documents collections of environmental data from the government and other sources. Potential users of the system include government agencies, industry, as well as the general public. Technically, the UDK is based on a three-way object model that distinguishes between environmental objects, environmental data objects, and UDK (meta) objects. Each (real-world) environmental object is described by a collection of environmental data objects. Each environmental data object is in turn associated with exactly one UDK object that specifies its format and contents. UDK objects are classified in accordance with an object-oriented inheritance hierarchy. They may be connected in a hypertext fashion to represent semantic associations, such as maintenance responsibilities and aggregation relationships. In January 1995, a first version of the UDK was made available in Austria; most German states will follow shortly. We also expect a prototypical version of the UDK to be available on the World Wide Web by early 1996.

Oliver Günther

Nick Van Driel and Tom Loveland

The U.S. Geological Survey's Land Cover Characterization Program

Abstract

The U. S. Geological Survey's (USGS) National Mapping Division has initiated a Land Cover Characterization Program in response to the need for land cover and vegetation data for inventory, monitoring, modeling, and management in the public and private sectors. The Program's general goal is a multi-scale, multi-purpose land characteristics data base. Customer needs, source data, and analytic techniques, form the basis of the program's four components. The small scale component uses satellite imagery acquired by the National Oceanic and Atmospheric Administration's (NOAA) advanced very high resolution radiometer (AVHRR). Multidate satellite imagery have been analyzed and combined with ancillary data to produce a multilevel, geographically referenced land cover data base for global change research. Landsat thematic mapper data are the source for the intermediate-scale component that delivers processed satellite data and ancillary data sets to cooperators. The USGS will synthesize a national land cover characterization from cooperators' results. Current activity in the large-scale component uses digital orthophotoquadrangles (DOQ) as a source for land cover and land use interpretations in selected urban areas for the National Water Quality Assessment (NAWQA) Program. A special projects component will accommodate cooperative activities in areas where standard products are not satisfactory. The USGS Land Cover Characterization Program is compatible with current concepts of government operations, the changing needs of the land use and land cover data users, and the technological tools with which the data are applied.

Historical Perspective

The USGS has a long heritage of leadership and innovation in land use and land cover mapping. The USGS Anderson system (Anderson, and others, 1976) defined the paradigm for land use and land cover mapping that has been the model both nationally and internationally for over 20 years. USGS scientists successfully applied these principles by mapping the United States. Recently, the USGS demonstrated the utility of multi- purpose land cover characteristics data bases, which build on the previous USGS efforts (Loveland, and others, 1991; Reed and others, 1994). The land characterization approach involves the interpretation of multi-resolution and multi-temporal data into flexible data bases describing landscape types, processes, and conditions. The results of the research phase have been broadly accepted within the national and international environmental assessment community (Steyaert, and others, 1994).

The land cover characterization program (LCCP) is founded on the premise that the nation's needs for land cover and land use data are diverse and increasingly sophisticated. The range of projects, programs, and organizations that use land cover data to meet their planning, management, development, and assessment objectives have expanded significantly. The reasons for this are numerous, and include the improved capabilities provided by geographic information systems (GIS), better and more data intensive analytic models, and increasing requirements for improved information for decision making.

Guiding Principles

The LCCP builds on the heritage and success of previous USGS land use and land cover programs and projects. It will be compatible with current concepts of government operations, the changing needs of the land use and land cover data users, and the technological tools with which the data are applied. The program is founded on the following guiding principles:

Land characteristics data bases that are compatible with previous USGS land use and land cover products and compatible with Federal Geographic Data Committee (FGDC) standards. This is essential so that there is continuity that allows analysis of land use and cover change.

Inter-agency partnerships. The USGS will actively solicit the involvement of other federal and state agencies in all aspects of the program, including funding, planning, mapping, and applications. This program will assist federal agencies in combining their agency efforts to characterize the landscape. No single organization can afford the cost of a national land characterization program.

There is no one product that satisfies all users or their applications. Therefore, the program will be based on a flexible data base strategy that facilitates both standardized categorized land use, land cover, and land characteristics describing landscape processes and dynamics.

Multi-resolution products will be designed for use at the local, regional, and national levels.

User driven and implemented on a cooperative basis. This will assure allocation of efforts to pressing national priorities.

Provide cyclic data that permit analysis of landscape change.

The LCCP is based on an integrated strategy of research, applications, production, and data management.

Project Components

One land cover product for the Nation would simply be inadequate. The LCCP thus consists of four components, each with unique but complementary land cover, land use, and ancillary data products which are:

1.Small scale (1:2,000,000) national and global land cover characteristics data produced periodically from coarse resolution (e.g., 1-km) remotely sensed data.

2.Intermediate scale (1:100,000) national land cover characteristics data produced on a cyclic basis from intermediate scale remotely sensed data such as data from the Landsat thematic mapper (TM).

3.Large scale (1:24,000) land use/land cover data for metropolitan areas and other areas experiencing rapid growth. Generally, this component will use the digital orthophotoquadrangle data as a mapping source.

4.Special projects will be conducted, as needed, to produce unique land cover characteristics data in situations in which the first three standard products do not satisfy user requirements.

As stated in the guiding principles, all four components will be developed through partnerships that include public and private participants. The following sections summarize the key aspects of each component. Note, however, that product definitions, timing, and other specific features will be thoroughly reviewed and refined during the first year of the program.

Small - Scale Component

The small - scale product will be a land cover characteristics raster data base with a grid cell size of 1-km2. The product will be designed for use in ongoing national operational programs with requirements for coarse resolution, nationally consistent data. This data set will be produced on a ten-year cycle corresponding to national population census periods. An experimental 1990 data base produced by the USGS is currently available (Loveland and others, 1993). The next national update will be in the year 2000.

Initially, the data set will be produced from advanced very high resolution radiometer (AVHRR) satellite data. Data from advanced sensors, such as the Moderate resolution imaging spectrometer (MODIS) may be used in the future. This product includes:

Seasonal land cover regions as the spatial component. They represent common mosaics of land cover, phenology, and landscape biophysical processes. See Loveland and others, 1995 for a description of seasonal land cover regions.

Attributes describing each region, including land cover types, vegetation components, phenology, spectral measures, and site characteristics (political boundaries, hydrology, elevation, general soils, climate, and ecoregions).

Derived thematic maps that are based on a translation of the seasonal land cover regions into common land cover classification legends; these include USGS Anderson System (Anderson and others, 1976), National Terrestrial Land Cover (Jennings, 1995), Biosphere-Atmosphere Transfer scheme (Dickenson, 1986), and Simple Biosphere Model (Sellers and others, 1986).

Source data used in the development of the data base, including all non proprietary satellite data and related earth science data sets.

A global land cover characteristics data base will be also produced as part of the LCCP. This effort is underway; a global 1-km data base produced from 1992-1993 AVHRR data will be completed by 1997. This product is intended for use in continental to global scale studies.

Intermediate-Scale Component

This product will be the primary detailed national land cover product. The data base, developed from Landsat TM satellite data or an equivalent source, will be a raster product with 30-m resolution. The data set is intended for use in regional (e.g., state, multi-county, ecoregions) land management, planning, and environmental assessment applications. As currently planned, this component will be implemented in two phases:

Phase 1: Every five years, a national preprocessed Landsat TM data set will be released by the USGS to United States Government cooperators. This data set will be georeferenced and will include appropriate ancillary data.

Phase 2: Every 10 years, the USGS will synthesize a national land cover characteristics data base. The process for the generation of this product will be based on interpretations made from a consistent, preprocessed Landsat TM data base. The primary source of these interpretations will be projects conducted by state and federal agencies, and the private sector. For example, the National Biological Service Gap analysis program, with state-level land cover mapping projects in most states, will be a key source of project-level land cover data. While the USGS will not dictate a standardized land cover legend for these projects, a set of specifications will be developed regarding minimum documentation standards. To facilitate access to the project-level data sets, the USGS will maintain an archive of land cover products developed from the national TM data base. As national land cover interpretations are completed, the USGS will synthesize the project-level interpretations into a nationally consistent classification legend, add appropriate attributes, and distribute the data base to anyone. This component is based on the current interagency land cover mapping initiative coordinated by the Multi-Resolution Land Characterization (MRLC) consortium (Loveland and Shaw, 1995).

Large-scale component

The large-scale component will emphasize land use and land cover. Keeping with the approach developed by Anderson and others (1976), land cover will be used as a surrogate for land use. The large scale data set will be vector format and will be developed from the interpretation of digital orthophoto quadrangle data. The data set will be consistent with the 1:24,000-scale USGS quadrangle maps. Unless special circumstances arise, source data will be used that can be released with the land use and land cover interpretations. This data set will be produced for metropolitan areas and other rapidly growing parts of the country. The target is to produce land use and land cover data every 10 years for all Standard Metropolitan Statistical Areas. The classification legend will be similar to the USGS Anderson System so that comparisons with the previous USGS land use and land cover data set are possible. However, refinement of the Anderson System is likely. The LCCP will work closely with the proposed FGDC Land Cover Subcommittee and with the NAWQA Program to define the final classification legend.

Special Projects

In cases where standard products will not satisfy user requirements, cooperative activities will be initiated to generate the necessary products. The modes in which special projects are conducted will be determined on a case-by-case basis.

Coordination

Many organizations and individuals conduct land cover and vegetation mapping, often without knowledge of each other's work. However, many of these activities are known to analysts and managers in the LCCP, and several initiatives are underway to provide coordination at state, regional, and national levels. Specifically, the Department of the Interior has created a Land Cover Working Group under the Interior Geographic Data Committee (IGDC). This working group, with membership from all DOI bureaus, will gather and publish information on each bureau's land cover and vegetation mapping plans, current activities, and related information sources. This information will be made available through the World Wide Web. In FY97, the IGDC Land Cover Working Group plans to expand its membership to other agencies under the FGDC. Because the FGDC includes 14 agencies, this working group will provide a coordination forum for all federal land cover and vegetation mapping activities. Some of these federal activities include coordination components at the national, regional, and state levels; these components will be used and expanded where appropriate. In addition, the LCCP will follow the successful MRLC Program example of coordination at the project level among agency programs with requirements for land cover data. MRLC participants include the USGS, NBS, EPA, NOAA, and USFS.

Outreach Activities

Communication of the LCCP's activities and goals will be the program's outreach focus in fiscal year 96. Information will be delivered through existing land cover and vegetation mapping projects and cooperatives, on the WWW under the USGS National Spatial Data Infrastructure node, through committees at bureau, department and agency levels, and by presentations at professional meetings and conferences. These activities will announce the program's products and schedules, and will encourage partnerships among individuals and groups with similar goals.

Standards

The USGS will work with the FGDC to establish a land cover working group. The initial charges of the working group will be to share information about federal plans and activities, develop a national land cover legend for the intermediate scale synthesized product, assist the FGDC land cover classification subcommittee and to write documentation standards for federal land cover projects.

The USGS will lead an effort to develop a crosswalk or conversion process for existing land cover classifications. With this conversion capability, valuable maps and data can be used as historic references in landscape monitoring activities.

Conclusion

Land cover, land use, and vegetation mapping data are essential elements in a wide array of government and private sector activities, including inventory, management, monitoring, and modeling. The time and expense required to develop land cover and vegetation data demand efficient collection and analysis procedures that allow multiple use of a single product. The LCCP reflects the U.S. Geological Survey's commitment to providing high quality land cover data for effective stewardship of the nation's resources.

REFERENCES

Anderson, J.R., Hardy, E.E., Roach J.T., and Witmer R.E. 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. U.S. Geological Survey Professional Paper 964, Reston, VA: U.S. Geological Survey.

Dickinson, R.E., A. Henderson Sellers, P.J. Kennedy, and M.F. Wilson. 1986. Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model. NCAR Technical Note NCAR/TN - 275+STR, Boulder, CO.

Jennings, M.D., 1995. Nomenclature and mapping units for gap analysis land cover data, in Technologies for Biodiversity Gap Analysis:Proceedings of the ASPRS/GAP Symposium, Charlotte, NC (in press).

Loveland, T.R., Merchant, J.W., Ohlen, D.O., and Brown, J.F., 1991. Development of a Land Cover Characteristics Data Base for the Conterminous U.S., Photogrammetric Engineering and Remote Sensing, v. 57, n. 11, p. 1453 - 1463.

Loveland, T.R., Ohlen, D.O., Brown, J.F., Reed, B.C., Merchant, J.W., and Steyaert, L.T., 1993. Prototype 1990 Conterminous United States Land Cover Characteristics Data Set CD-ROM, USGS CD-ROM Set 9307.

Loveland, T.R., and Shaw, D.M., 1995. Multiresolution Land Characterization:Building Collaborative Partnerships, in Technologies for Biodiversity Gap Analysis: Proceedings of the ASPRS/GAP Symposium, Charlotte, NC (in press).

Reed, B.C., Loveland, T.R., Steyaert, L.T., Brown, J.F., Merchant, J.W., and Ohlen, D.O., 1994. Designing Global Land Cover Databases to Maximize Utility, in Environmental Information Management and Analysis: Ecosystem to Global Scales, W.K. Michener, J.W. Brunt, and S.G. Stafford, editors: London, Francis and Taylor, p. 299 - 314.

Sellers, P.J., Y. Mintz, Y.C. Sud, and A. Dalcher. 1986. A simple biosphere model (SiB) for use within general circulation models. Journal of Atmospheric Science 43: 505-531.

Steyaert, L.T., Loveland, T.R., Brown, J.F., and Reed, B.C., 1994. Integration of Environmental Simulation Models with Satellite Remote Sensing and Geographic Information Systems Technologies: Case Studies, in Proceedings:Pecora 12 Symposium on Land Information from Space-Based Systems, American Society of Photogrammetry and Remote Sensing, Bethesda, MD, p. 407 - 417.

Nick Van Driel
Tom Loveland
EROS Data Center
Sioux Falls South Dakota, 57198 Phone: 605-594-6007
FAX: 605-594-6529

Land Use History of North America- Need for a Continental Synthesis

Background: The value of a land use history

Efforts to manage the nation's biological resources are hampered by the lack of an historical perspective on conditions prior to European settlement and subsequent changes in the North American landscape. Much of the impact that people have had on the environment can be viewed as a series of unplanned experiments, with particular perturbations generating measurable responses, in the form of contractions in the ranges of some species and expansions in the ranges of others. Within the context of these temporal dynamics, species extinctions and the spread of non-indigenous species may be seen as the extreme cases, where biological elements are lost or introduced. These experiments have been run, and environmental scientists are beginning to assemble the data needed to assess the results. The first task is to develop a clearer understanding of the historic changes in the distributions of plants and animals and their relation to human- induced changes to the landscape. Given this understanding, land managers will be able to review the effects of past perturbations and apply this information when attempting to evaluate the likely outcomes of future land changes.

Much of the data needed to construct such a retrospective view have already been collected; information on landscape change span the period of human habitation of North America. Impressive regional efforts have been undertaken to synthesize the available information regarding land use change and its impact on ecological systems, but these projects have generally been limited to relatively small areas and short time lines. Large quantities of valuable biological and physical information remain unexplored, warehoused in different locations, and maintained by different organizations.

Consider the abundant information on prehistoric land cover and species distributions accumulated through the creative efforts of paleoecologists. Integration of coarse-resolution data such as these with information derived from original land surveys of the country (e.g. data archived by the Bureau of Land Management), and the U.S. Forest Service's data on the fire history of North America, for example, could make the characterization of historic landscape change quite tractable. When these data are combined with aerial photography from the extensive surveys started in the 1930's, and remotely sensed data from advanced satellite imagery, it will be possible to stitch together a continuous time line, from prehistoric times to the present. Catalyzing such an effort is the intent of the National Biological Service's Land Use History of North America project (LUHNA.)

Launching the Project

This is an ambitious project, one that will require the collaboration of many different individuals and agencies, both within and outside government. In August, the National Biological Service (NBS) convened a workshop to help define the scope and intent of the LUHNA effort, and to identify a strategy for fostering the multidisciplinary collaboration that it will require. Representatives from six government agencies, six universities, and three not-for-profit organizations established a framework for building a broader LUHNA effort. NBS will serve as the organizer and initial "home base" for the project, but NBS cannot possibly carry out such a large project alone, nor fund all the work that will be required. Instead, NBS will provide a forum for discussion, communication, and scoping of the project and the initiative for developing the interdisciplinary relationships that LUHNA will require. NBS will also approach other organizations for cooperation and/or funding support.

NBS is maintaining information on development of LUHNA on and Folio Infobase server accessible through the NBS Homepage (http://www.nbs.gov) on the World Wide Web. Select NBS Infobase Server and click on LUHNA. Your email comments are invited at luhna@ibis.mib.nbs.gov.

Carol A. Johnston, Carol A. Sersland, John Bonde, Deb Pomroy-Petry, and Paul Meysembourg

CONSTRUCTING DETAILED VEGETATION DATABASES FROM FIELD DATA AND AIRBORNE VIDEOGRAPHY

ABSTRACT

Ecologists require new methods for constructing detailed, spatially-explicit vegetation databases. Several methods were evaluated: (1) statistical classification of field data, (2) statistical ordination of field data, and (3) airborne videography. Plant species cover was measured at 81 field sample points distributed throughout a 2.9 ha wetland. The location of each sample point was determined using a Motorola LGT 1000 GPS with real-time differential processing. Plant communities were classified from the cover data using two-way indicator species analysis (TWINSPAN), a classification technique, and Canonical Correspondence Analysis (CCA), a multi-variate ordination technique. Thiessen polygon maps of plant community distribution were generated from the TWINSPAN analysis using an ARC/INFO geographic information system, and color composite maps were generated from the CCA analysis using an ARC/INFO geographic information system and an ERDAS Imagine image analysis system.

Color aerial video imagery of the same wetland was obtained at two different phenological stages, summer and fall of 1994, and at varying altitudes. Frames were georeferenced, rectified, mosaicked, and composited into a multi-temporal image. Vegetation was classified using a supervised classification routine. Comparison of the classified video image with the Thiessen polygon maps showed a 60% correspondence. Although the maps generated using field data provided more detailed classification of plant community types, the airborne video maps provided a more detailed spatial depiction of their distribution.

INTRODUCTION

Ecologists have become increasingly interested in spatial relationships among plant communities. Traditional approaches to the evaluation of such relationships have included statistical evaluation of data collected at sample points, and classification of vegetation from satellite imagery. Scales of measurement for the first approach are square centimeters to meters, whereas scales of measurement for satellite imagery are 100 to 1,000,000 square meters. Techniques have been lacking for intermediate scales.

Recent advances in GIS, GPS, interpolation techniques, multivariate statistics, and airborne remote sensing are providing ecologists with new tools for constructing spatial databases of vegetation. We classified and mapped wetland vegetation using multi- temporal airborne videography and multivariate statistical analysis of field data. A global positioning system (GPS) was used to georeference both data sets. Image analysis and GIS interpolation techniques were used to prepare maps from the georeferenced airborne videography and field data sets, respectively.

METHODS

Study Site

A 2.9 ha wetland along the Pokegama River near Superior, Wisconsin, USA (46o40'35" N, 92o8'51" W) was selected for study. The site consists of wetlands in the river bed, a low natural levee, and wetlands in the backwater behind the levee. Trees, shrubs, and herbaceous vegetation grow on the levee, whereas only herbaceous vegetation grows in the river bed and backwater areas.

Sample points were established in transects 20 m apart, oriented perpendicular to the river using a compass and fiberglass tape measure. Each transect consisted of a sample point in the riverbed, on the levee, and at 2-5 locations in the backwater. Sample points were spaced ~20 m apart along the transects in backwater areas, but were closer (4-10 m) in the riverbed and on the levee, where environmental gradients changed rapidly. A total of 81 sample points were established, and marked with steel rods.

Georeferencing

The location of each sample point was determined with real-time differential processing (Hurn 1989) using two Motorola LGT 1000 global positioning systems (GPS). One LGT 1000 was used as a portable base station by positioning it over a National Geodetic Survey (NGS) benchmark overlooking the Pokegama River valley. Latitude, longitude, and elevation specifications obtained from NGS for the benchmark were programmed into the base station GPS. This GPS was linked to a UHF radio, which transmitted instantaneous correction data every 5 seconds to the field unit. Power was provided by two 7 amp-hour 12VDC utility batteries, enough to run the GPS and radio transmitter for about 10 hours. The GPS, radio, and battery fit into a 10"x10"x10" wooden box for convenient carrying. The same base station equipment was used to provide real-time differential correction for the GPS-linked airborne video (see below).

The Motorola LGT 1000 GPS used as the field unit was connected to a hand-held UHF radio, which received the correction data and provided voice communication with the base station. Both pieces of equipment were mounted onto a back pack frame for portability. The detachable GPS antenna was mounted onto a 2 foot PVC pole, which was positioned over the steel rods marking each sample point, and which could be elevated above any tall herbaceous vegetation which might obscure satellite signal reception (Fig. 1). Readings were taken for approximately 1.5 minutes per location, yielding a minimum of 100 GPS readings that were averaged to obtain the final coordinates.

Vegetation and Environmental Field Data

Plant cover by species was measured in August 1993 using the line intercept method (Mueller-Dombois and Ellenberg 1974). Cover at each sample point (referred to hereafter as a "plot") was recorded over one meter of line oriented parallel to the river, and over a second meter of line oriented perpendicular to the first, because of the potential for anisotropy. All intercepting species were recorded, but where two or more species overlapped, only the dominant (overstory) layer was measured, such that total cover never exceeded 100% of the line lengths measured. Taxonomic keys (Fassett 1957, Britton and Brown 1970) and pressed specimens from the Olga Lakela Herbarium at the University of Minnesota, Duluth were used in plant identification. Six environmental variables were also measured for each sample point: water depth, soil organic matter, sand content (0.05 to 2.0 mm particles), silt content (0.002 to 0.05 mm particles), clay content (< 0.002 mm particles), and soil pH.

Multivariate Statistical Analysis of Field Data

There are two basic conceptual models for statistically analyzing matrices of ecological data (Pielou 1984). "Classification" arranges sample points into discrete groups, such as plant communities. "Ordination" arranges sample points and/or species along environmental gradients. We used examples of both to analyze the field data collected: two-way indicator species analysis (TWINSPAN) is a classification technique, and Canonical Correspondence Analysis (CCA) is an ordination technique. FORTRAN computer programs are available in the public domain for performing these analyses (Hill 1979, ter Braak 1991).

Two-way indicator species analysis (TWINSPAN) is a divisive classification technique devised to partition reciprocal averaging (RA) ordinations (Pielou 1984). It treats all plots as a single entity at the onset, and iteratively divides them into hierarchical groups. TWINSPAN performs a one-dimensional RA ordination and breaks the axis at the centroid so as to crudely divide the data points into two classes; this procedure is repeated with the species quantities weighted in such a way as to emphasize the influence of especially useful diagnostic species (i.e. "indicator" species) identified by the first ordination. TWINSPAN produces a dendrogram showing relationships between plot groupings, such that groups positioned close to each other in the dendrogram are most similar. Only vegetation data were used in the TWINSPAN analysis performed here.

Canonical Correspondence Analysis utilizes both vegetation and environmental data. CCA is a direct gradient analysis technique, in which species composition is directly and immediately related to measured environmental variables. It mathe-matically defines canonical axes, and produces diagrams of sample points and plant species plotted in relation to those axes. The Cartesian coordinates of each sample point relative to the first three canonical axes were used in subsequent GIS analysis. The mechanics and advantages of CCA over other ordination techniques are described in detail by Palmer (1993).

It should be stressed that neither TWINSPAN nor CCA utilize spatial statistics, so the location of each sample has no bearing on its statistical treatment. Each program mathematically synthesizes information about vegetation, and describes species assemblages at each point in space that was sampled. Also, neither program produces vegetation maps; the interpolation capabilities of a GIS are required to generate maps from these point data.

GIS Analyses

The two multivariate statistical methods yielded different types of output, so different GIS methods were needed to construct maps from their results. TWINSPAN assigned each sample point to a discrete vegetation class, suitable for GIS analyses involving categorical (nominal) data. These data were used to generate Thiessen polygon maps (Green and Simpson 1978) with the ARC/INFO Geographic Information System. Boundaries between Thiessen polygons of like vegetation class were dissolved in ARC/INFO. A data layer of the wetland boundary was generated by displaying the sample points on the computer screen and circumscribing the boundary based on its appearance on the airborne video image (Fig. 2); this data layer was used to clip off Thiessen polygons that extended past the wetland boundaries.

CCA generates canonical axis values for each sample point. As with any ordination technique, these are unitless numbers which may be positive or negative. Large positive numbers indicate sample points with a strong positive relationship to the CCA axis, and large negative numbers indicate those with a strong negative relationship. Point data layers were constructed for each of the first three CCA axes by assigning canonical axis values to their corresponding sample point locations. These point values were then interpolated using an inverse squared distance weighted moving average (Burrough 1986) with the IDW command in ARC/INFO GRID, resulting in grid of cells. The interpolation was performed using the 12 closest sample points, with the wetland boundary used as a barrier to limit the search for input points. The output cell size was 5 x 5 meters. The grids were converted to ERDAS Imagine files, and these files were joined in Imagine using the "stacklayer" command to create a three-band image. Each of the three bands was assigned a different color (red, blue, green) for simultaneous display. In the resulting color composite map, different color combinations indicated different combinations of canonical axis values.

Airborne Videography

Video imagery was acquired in July and September 1994 from a Cessna Skyhawk aircraft at flight altitudes of 2000 and 1000 feet, respectively. Flights took place between 10 am and 2 pm to minimize shadow length. On-site observations were made within two days of the actual flight to account for phenological variables such as plant growth stage and background characteristics such as turbidity and amount of open water. Styrofoam targets were placed in areas of homogeneous wetland vegetation for use in supervised training procedures during subsequent image processing and to provide ground control data for verifying the video georeferencing (Sersland et al., 1995). The ground location of each of the eight targets was determined by processing carrier phase GPS information collected by Motorola LGT1000 receivers.

Aerial video imagery was recorded using a Panasonic GP-KR412 VHS color composite video camera head with a 12 mm focal length. This camera senses in the visible portion of the electromagnetic spectrum, from 0.42 - 0.85 um, though the exact spectral response in the red, green, and blue has not been determined. The video signal was recorded on 1/2 inch video tape in a Panasonic AG-7355 S-VHS video tape recorder connected to the camera head.

A Motorola LGT1000 Global Positioning System (GPS), programmed to receive a locational fix once a second, was integrated with the aerial video camera system on the July flight. Real-time differential corrections were provided by another LGT1000 unit on the ground using a UHF radio link. The Motorola was replaced by a 24-channel Ashtech 3DF ADU receiver on the September flight. This receiver computed the plane's position twice a second, and determined its attitude (roll, pitch and yaw) from four GPS antennas mounted on each wing and the front and rear of the fuselage. A data encoding system was installed in the VCR that interfaces with the GPS receiver to frame register the video data.

The in-plane system was controlled by a notebook computer. For each GPS location, the current SMPTE code (unique identifier) was queried from the video recorder. The latency in the GPS data was adjusted for and the result was a data file containing plane position and attitude information for two video frames each second.

In the lab, a 486 PC with a Truevision Targa+ frame-grabber connected to the Panasonic AG-7355 VCR automatically positioned the video tape to a given frame by entering the frame's SMPTE code and then rasterizing the image. After the selected images were frame- grabbed, they were automatically rectified, converted to a common scale, and mosaicked together into one georeferenced file using software written for PC ERDAS 7.5. The June and September images were co-registered to make a multi-date composite, using the targets and other identifiable objects for ground control. The two images were then layered using the routine LAYERSTACK in Imagine to produce a six band image. Training signatures were defined in the Signiture Editor with ERDAS Imagine's Area of Interest/Seed Properties, and merged until the maximum separation between classes was achieved. A supervised classification with the maximum likelihood decision rule was applied. The vegetation field data collected in August 1993 (see above) were used to evaluate classification accuracy. Details of this portion of the study are described by Sersland et al. (1995).

RESULTS

Georeferencing

The Motorola LGT1000 GPS units were versatile, and worked well in field and airborne applications under either carrier phase or real-time differential mode. Carrier phase readings provided greatest accuracy, within 3 cm of true ground location at the 95% probability level. Such accuracy was needed for georeferencing the targets for video ground truth because of the sub-meter resolution of the video data. The use of carrier phase was slow, however, requiring that the units be left stationary for about 20 minutes at each ground location, which would have been prohibitively long for georeferencing the 81 vegetation sample points.

Real-time differential readings using conventional (C/A) GPS codes provided accuracies consistent with the spatial scale of the study, within 2-3 m of true ground location at the 95% probability level. If we had not used differential processing, the GPS readings would have been accurate only to within 100 m at the 95% probability level. The use of real-time differential processing overcame a potential equipment limitation caused by the steep terrain of the Pokegama River valley: given that the 6-channel Motorola LGT1000 GPS units could each receive signals from only six satellites, there was the potential that the base unit could have been receiving signals from satellites blocked from the field unit's view by the steep valley sides. Real-time differential processing allowed us to make sure that the same satellite signals were being received by both the field and base units, and that the GPS data quality was adequate prior to leaving the field site.

The use of real-time differential processing also aided field and aerial navigation. A number of the steel rods marking field sample points were submerged by high water levels during the spring after they were installed, but we were able to retrieve them by using real- time differential GPS to navigate to the coordinates we had measured the previous field season. In the airplane used for video image acquisition, the coordinates of target sites were pre-programmed into the airborne GPS to serve as "waypoints" to guide navigation, providing the pilot with information about the distance and direction to the target areas. Without real-time differential correction, the GPS readings would have been too imprecise for navigation.

The Ashtech 3DF ADU is a more expensive and specialized GPS, designed to provide attitude data in real-time to accuracies of about 0.1 degrees with a 2.5 meter antenna separation. These attitude data were used to correct video image georeferencing errors induced by aircraft roll, pitch, and yaw (Fig. 2). Post-processed or real-time accuracies were within 5 meters for each position computed (up to twice a second).

Although field workers were directed to set up the field sampling locations on an evenly-spaced grid within the backwater area, the GPS data revealed a large gap in the sampling grid behind a convex curve of the levee (Fig. 2). If we had assumed that the grid was rectilinear rather than determining GPS locations for each of the sample points, the vegetation maps resulting from the field data would have been distorted in this region. GPS georeferencing also allowed us to interface the field maps with other georeferenced data, such as National Wetland Inventory maps, which would not have been possible using coordinates determined by tape measure.

TWINSPAN/Thiessen Polygon Map

Seven plot groupings were distinguished by six iterations of TWINSPAN, and named for the one or two species that were most abundant within each group. The Alnu covertype is dominated by green alder shrubs (*Alnus crispa*), CarS is dominated by tussock sedge (*Carex stricta*), CaTy is a mixture of sawgrass sedge (*Carex lacustris*) and cattail (*Typha latifolia*), Frax is dominated by black ash trees (*Fraxinus nigra*), SagG is dominated by arrowhead (*Sagittaria graminea*) and associated aquatics, Scir is a monospecific stand of three-square rush (*Scirpus americanus*), and Spar is dominated by large-fruited burreed (*Sparganium eurycarpum*). The Alnu and Scir classes consisted of only one plot each.

The Thiessen polygon map generated from these data appeared somewhat blocky, but revealed the major spatial patterns of vegetation distribution (Fig. 3). Wetlands growing in the bed of the Pokegama River were dominated by the Spar covertype. Spar stands were also common in backwater areas of intermediate water depth. SagG and CaTy occurred as large, contiguous polygons in the deepest and shallowest portions of the backwater, respectively. The levee consisted of eight plots classified as CarS, four plots classified as Frax, and one plot classified as Alnu. The single Scir plot occurred in the riverbed near the northernmost tip of the study site.

The sizes and shapes of each Thiessen polygon were determined by the proximity and location of adjacent sample points. Thiessen polygons define individual areas of influence around each point in such a way that the polygon boundaries are equidistant from neighboring points, and each location within a polygon is closer to its contained point than to any other point (Maggio and Long 1991). Polygon boundaries were close to the sample points where the points were closely spaced, such as between the riverbed and the levee, but far from the sample points in the area behind the convex curve of the levee where there was a gap in the sampling grid (Fig. 3). Each covertype classification is correct at the point where it was measured, but classification certainty decreases with distance from that point. The Thiessen polygons are therefore not necessarily representative of the true spatial extent of vegetation types, particularly in areas where sample points are widely spaced.

CCA/Color Composite Map

The first three canonical axes were primarily related to water depth, soil organic matter content, and soil pH, respectively. When the interpolated data layers for each of the three axes were displayed simultaneously as a three-band image, the result was a confusing array of colors that was difficult to interpret (Fig. 4a). Therefore, the data layer for the third canonical axis, which did not substantially increase the cumulative percentage variance explained by the first two canonical axes, was dropped from further analyses.

The first canonical axis separated the levee from the deep backwater, and the second canonical axis separated the low organic matter soils from high organic matter soils. The color composite map for these two CCA axes illustrates these interactions (Fig. 4b). Values for CCA axis 2 were assigned to the red layer for display. Values for CCA axis 1 were assigned to the blue and green layers, which combine to make cyan. Neutrals (whites and grays) on the composite map indicate areas where the values for CCA axes 1 and 2 are similar, white indicating an area where both values are high, and dark gray indicating an area where both values are low. Highly saturated colors (ie. bright red, cyan) indicate areas where the values for the CCA axes are opposite: where one is high, the other is low. Intermediate colors represent intermediate values for each CCA axis.

Colors on the composite map for CCA axes 1 and 2 (Fig. 4b) can be interpreted as follows:
light pink, light cyan = no surface water, low to intermediate organic matter soils (ie. levee)
dark gray = deep surface water, high organic matter soils
bright red = intermediate to deep water, low organic matter soils (ie. riverbed, northern portion of backwater)
bright cyan = shallow surface water, high organic matter soils (ie. southern portion of backwater)
dark red, dark cyan = intermediate water depth and soil organic matter (ie. central portion of backwater)

The first two CCA axes explained 74.7% of the cumulative variance of species-environment relationships, but only explained 11.6% of the cumulative variance of the plant species data alone, because many of the species were widely distributed across the environmental gradients represented by the CCA axes. The distribution of species such as Carex stricta and Carex lacustris were well explained by CCA axes 1 and 2, respectively, but the distribution of more ubiquitous species (e.g. Sparganium eurycarpum) was poorly explained. Therefore, although this map depicts environmental gradients well, it does not depict species distribution as well. This would not be the case in an ecosystem in which plant communities have discrete species assemblages and environmental requirements.

Airborne Videography Map
Nine cover classes were developed from the image analysis and accompanying field work (Table 1). Burreed (BU) and cattail-sedge (CS) were the most abundant covertypes, each constituting 31% of the study site area (Sersland et al. 1995). Other vegetation types of large extent were cattail-burreed (CB) and arrowhead (AC, AT). Only small areas of tussock sedge (TS) and trees & shrubs (TR) were identified (Figure 5).

Table 1. Classes of vegetation distinguished by image analysis, with equivalent TWINSPAN classes.

			TWINSPAN

	CODE DESCRIPTION	DOMINANT PLANT SPECIES	EQUIVALENT
RI	River	--	none
BU	Burreed	*Sparganium eurycarpum*	Spar
TR	Trees & shrubs	*Fraxinus nigra, Alnus crispa*	Frax, Alnu
TS	Tussock sedge	*Carex stricta*	CarS
AC	Arrowhead, clear water background	*Sagittaria graminea*	SagG
AT	Arrowhead, turbid water background	*Sagittaria graminea*	SagG
CS	Cattail-sedge	*Typha latifolia, Carex lacustris*	CaTy
CB	Cattail-burreed	*T. latifolia, S. eurycarpum*	Spar
SA	Submersed aquatic	*Nuphar variegatum, Ceratophyllum*	none

Classification accuracy, determined by comparing the August 1993 field data with the classified multi-date composite image, was about 60% (Sersland et al. 1995). Previous attempts to use conventional color videography for mapping wetland vegetation have been less successful (Jennings et al. 1992). The level of accuracy we achieved by classifying a multi-date composite color image was comparable to the levels of accuracy obtained using more expensive multi-spectral videography (Bartz et al. 1992, Thomasson et al. 1994).

The levee vegetation was particularly poorly identified (Figure 5). Several factors may contribute to this: (1) narrow configuration - in many places, the levee was only as wide as a single tree canopy, (2) leaf-off imagery - the September image was acquired after the leaves had dropped from the trees and shrubs, which would have made them difficult to detect (3) high biodiversity - levee vegetation consisted of a diverse assemblage of species having different spectral reflectance qualities, (4) shadows - the shadows cast by levee trees and shrubs on the July image could have confused the classification. Classification accuracy was improved to 66% correct by excluding reference data points vegetated by trees & shrubs (Sersland et al. in press).

There was good correspondence (52%) between the map generated by image classification and the map generated by Thiessen polygons from TWINSPAN analysis of the field data (Sersland et al., 1995). The two maps corresponded well in the extensive cattail-sedge area in the southern portion of the study site, and in burreed beds scattered throughout the study site. Arrowhead beds also corresponded well in the northern portion of the backwater area.

Both the image map and the Thiessen polygon map have inherent limitations, so one cannot be thought of as more "correct" than the other. The Thiessen polygon map has high classification accuracy at the center point of each polygon, but uncertain classification accuracy elsewhere. The image-derived map has high classification accuracy for certain covertypes, but low classification accuracy for others. The spatial resolution of the airborne video image was much smaller (pixel size = 1 m2) than the spatial resolution of either of the maps generated from field data. As a result, the minimum mapping unit is much smaller, and boundary configurations between vegetation categories appear much more natural (Figure 5).

DISCUSSION

Vegetation mapping is greatly affected by the methods used to measure, classify, and discretize floristic data. Each of the methods examined yielded different results, despite the common study site and, in the case of the TWINSPAN and CCA analyses, the common data set.

The use of field data to map vegetation is preferred by most botanists, because they are able to count and measure individual species. Remote sensing is by definition a "hands-off" technique, which is less satisfying to scientists accustomed to studies at the scale of individual species. Remote sensing techniques are less discriminating of background information, such as water turbidity, which could be totally screened out by a field botanist. Given that this background information may be ecologically significant, its detection is not necessarily bad, but it complicates preparation of vegetation maps.

Because floristic data are complex, with many species co-occuring at the same location, analysis and/or interpretation is required in order to classify the data into vegetation assemblages. A traditional approach has been to perform this classification in the field, with the botanist assigning vegetation into pre-determined community types. This approach works well where vegetation occurs naturally in distinct assemblages, but is less suitable where vegetation distributions overlap each other. Statistical methods such as TWINSPAN and CCA provide new tools for distinguishing vegetation assemblages more objectively.

Point data obtained by field measurements are 100% correct at the points where they are sampled, assuming that there is no measurement, classification, or recording error. Constructing a map, however, requires that point data be interpolated. Uncertainty increases with distance from the sample point, so any interpolation technique introduces error. This error is greatest in areas of low sampling density, and in areas where vegetation characteristics change very quickly over space, as in the vicinity of the levee at our study site.

Constructing a map from point data requires collection of locational as well as floristic data, a fact often overlooked by field ecologists. Prior to the development of GPSs, accurate locational data had to be acquired by expensive surveying methods, which required bulky equipment and the presence of a surveyed benchmark in the vicinity. This made conventional surveying very impractical in remote and marshy areas. The development of GPS and the decreasing costs of GPS equipment have made it an indispensible tool for spatial ecology.

Vegetation maps are traditionally categorical, depicting plant groupings as large, discrete entities delimited by distinct boundaries (Fig. 3). Maps derived from remote sensing are also categorical, but the minimum mapping unit is constrained not by sampling density, but by the spatial resolution of the sensor. In the case of our airborne videography, the minimum mapping unit was much smaller and the number of polygons was much larger than that of the more conventional Thiessen polygon map (Figs. 3, 5). Vegetation maps derived from continuous numerical data are more difficult to interpret than categorical maps, but may depict vegetation and environmental gradients more faithfully than do categorical maps (Fig. 4).

In summary, all of the techniques we used produced useful vegetation maps, each with inherent inaccuracies. The technique of choice would depend upon the planned use of a particular map. The TWINSPAN/Thiessen Polygon Map would be preferable for the schematic depiction of discrete vegetation assemblages, the CCA/Color Composite Map would be preferable for depicting vegetation patterns relative to environmental gradients, and the Airborne Videography Map would be preferable for depicting fine spatial detail. Each map is merely a model of the complex reality that exists in nature.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge research support from the National Science Foundation (BIR-9208616), the USDA National Research Initiative Competitive Grants Program (USDA/92-37102-7406), and Minnesota Technology, Inc. Ann Lima performed the statistical analyses, and Jim Sales coded the manuscript into HTML for WWW publication.

REFERENCES

Bartz, K.L., J.L. Kershner, R.D. Ramsey, and C.M.U. Neale. 1992. Assessing Riparian Vegetation Using Multispectral, Airborne Videography: A New Resource Management Tool. Proceedings of the Fourth Forest Service Remote Sensing Applications Conference. American Society for Photogrammetry and Remote Sensing. pp. 319- 327.

Britton, N.L., and A. Brown. 1970. An Illustrated Flora of the Northern United States and Canada. Dover Publications, New York.

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press.

Fassett, N.C. 1957. A Manual of Aquatic Plants. University of Wisconsin Press, Madison.

Green, P.J., and R. Sibson. 1978. Computing Dirichlet tesselations in the plane. Comput. J. 21:168-173.

Hill, M.O. 1979. TWINSPAN - A FORTRAN Program for Arranging Multi- variate Data in an Ordered Two-Way Table by Classification of the Individuals and Attributes. Cornell University, Ithaca, NY.

Hurn, J. 1989. GPS: A Guide to the Next Utility. Trimble Navigation, Sunnyvale, CA.

Jennings, C. A., P. A. Vohs, and M. R. Dewey. 1992. Classification of a Wetland Area Along the Upper Mississippi River with Aerial Videography. Wetlands, Vol. 12,No. 3, pp. 163-170.

Maggio, R.C., and D.W. Long. 1991. Developing thematic maps from point sampling using Thiessen polygon analysis. pp. 1-10. In Proceedings, GIS/LIS'91, Atlanta, GA, Volume 1. American Society for Photogrammetry and Remote Sensing, Bethesda, MD.

Mueller-Dombois, D., and H. Ellenberg. 1974. Aims and Methods of Vegetation Ecology. John Wiley & Sons, New York.

Palmer, M.W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. Ecology 74:2215-2230.

Pielou, E.C. 1984. The Interpretation of Ecological Data. John Wiley & Sons, New York.

Sersland, C.A., C.A. Johnston and J. Bonde. 1995. Assessing wetland vegetation with GPS-linked color video image mosaics. pp. 53-62. In: Proceedings, 15th Biennial Workshop on Color Photography and Videography in Resource Assessment. American Society for Photogrammetry and Remote Sensing, Bethesda, MD.

ter Braak, C.J.F. 1991. CANOCO a FORTRAN program for community ordination by [partial][detrended][canonical] correspondence analysis, principal components analysis and redundancy analysis. Version 3.12. ITI-TNO, Wageningen, The Netherlands.

Thomasson, J. A. and C. W. Bennett, B. D. Jackson, and M. P. Mailander. 1994. Differentiating Bottomland Tree Species with Multispectral Videography. Photogrammetric Engineering and Remote Sensing, Vol. 60,No. 1, pp. 55-59.

Carol Johnston
GIS Administrator
Natural Resources Research Institute
University of Minnesota
5013 Miller Trunk Highway
Duluth MN 55811
Phone: 218-720-4269
FAX: 218-720-4219
Email: cjohnsto@sparkie.nrri.umn.edu

Carol Ann Sersland
Research Assistant
Department of Forest Resources
University of Minnesota
1530 North Cleveland Avenue
St.Paul, MN 55108-6112
Phone: 612-624-3459
Email: sersl001@maroon.tc.umn.edu

John Bonde
GIS Manager
Natural Resources Research Institute
University of Minnesota
5013 Miller Trunk Highway
Duluth MN 55811
Phone: 218-720-4266
FAX: 218-720-4219
Email: jbonde@sparkie.nrri.umn.edu

Deborah Pomroy-Petry

Research Plot Technician
Natural Resources Research Institute
University of Minnesota
5013 Miller Trunk Highway
Duluth MN 55811
Phone: 218-720-4357
FAX: 218-720-4219
Email: dpomroyp@sparkie.nrri.umn.edu

Paul Meysembourg
Principal Laboratory Technician
Natural Resources Research Institute
University of Minnesota
5013 Miller Trunk Highway
Duluth MN 55811
Phone: 218-720-4275
FAX: 218-720-4219
Email: paul@sparkie.nrri.umn.edu

James J. Majure, Noel Cressie, Dianne Cook, and Jürgen Symanzik

GIS, spatial statistical graphics, and forest health.

ABSTRACT

This paper discusses the use of a geographic information system (GIS), Arcview 2.1, linked with a dynamic graphics program, XGobi, in the statistical analysis of spatial data. The link allows multivariate data, collected at geographic locations and stored in Arcview, to be passed into XGobi and analyzed dynamically. The connection between the points in XGobi and the spatial locations from which they were collected is maintained so that points in either Arcview or XGobi can be brushed and the corresponding points in the other application identified immediately. Spatial cumulative distribution functions (SCDFs), spatially lagged scatter plots and variogram-cloud plots can be displayed in XGobi using the link. In each type of plot, the connection to the spatial sampling location is maintained and user interaction can take place in either application.

The link is used to predict and analyze SCDFs of forest crown health in the northeastern United States. The SCDFs are predicted from field data collected as part of the U.S. Environmental Protection Agency's (USEPA) Environmental Monitoring and Assessment Program (EMAP). The field data are augmented with concomitant geographic information, including Landsat Thematic Mapper images, digital elevation models, and population information, which are used to improve the SCDF prediction.

If you are interested, after reading this paper, click here to see documentation and download instructions for these tools.

INTRODUCTION

This paper discusses the integration of a dynamic graphics program, XGobi, into a geographic infomraiton system (GIS), Arcview 2.1 (ESRI 1995), and its use in the statistical analysis of spatial data. The link between XGobi and Arcview allows multivariate data, collected at geographic locations and stored in Arcview to be passed into XGobi and viewed. The connection between the points in XGobi and the spatial locations from which they were collected is maintained so that points in either XGobi or Arcview can be *brushed* (see Note 1 at the end of the paper), resulting in simultaneous brushing of corresponding points in the other application. The link also has the ability to use XGobi to display spatial cumulative distribution functions (SCDFs), spatially lagged scatter plots, and variogram-cloud plots. In each type of plot, the connection to the spatial sampling locations is maintained and user interaction can take place in either application.

The particular problem to which these tools are applied involves the prediction and analysis of SCDFs for forest crown health in the northeastern United States. The SCDFs are predicted from field data collected as part of the U.S. Environment Protectin Agency's Environmental Monitoring and Assessment Program (EMAP). The field data are augmented with concomitant geographic information, including Landsat Thematic Mapper images, digital elevation models, and population information, which are used to improve the SCDF prediction.

In this paper, we will first give an overview of the linking technology between Arcview and XGobi. We will then discuss the use of the link in the prediction of SCDFs.

INTEGRATION OF DYNAMIC GRAPHICS TOOLS INTO A GIS

Interactive and dynamic graphics programs are very useful in the exploration of high-dimensional data. With data collected at spatial locations, it is important to include the locations as part of the analysis. This leads very naturally to the integration of a GIS with a dynamic graphics program; the GIS is used for displaying spatial locations and concomitant geographic variables, and the dynamic graphics program is used for visualizing and exploring the corresponding data space. This type of link has been constructed between Arcview 2.1 and XGobi (Swayne et al. 1991), an interactive dynamic graphics program in the X Window SystemTM environment. Technical details of the link can be found in Symanzik et al. (1995) and Majure et al. (1995).

The link between Arcview and XGobi is intended to provide functionality that is not provided by either the GIS or the dynamic graphics program alone. While GISs provide sophisticated capabil ities for the input of spatial data, its management, and the display of maps, graphics and tables, their capability for statistical analysis is generally limited and dynamic graphical analysis is non existent. Although most dynamic graphics programs can plot the coordinates of spatial locations, they do not have the capabilities of producing high quality maps that provide a geographic frame of reference. Together, then, Arcview and XGobi share their strengths and produce a product that is more than the sum of the parts.

The specific tools made available by the link include the resident capabilities of both Arc view and XGobi, as well as the ability to do *linked brushing* (see Note 1 at the end of the paper) between the two systems. The capabilities of Arcview 2.1 include the display and manipulation of sample locations and other geographic information. XGobi provides an array of graphic options through the manipulation of scatter plots. The types of plots avail able include univariate and bivariate plots, three-dimensional point rotation, and higher-dimensional rotation with the grand tour (Asimov 1985, Buja and Asimov 1986) and the correlation tour (Buja et al. 1988). Both the grand tour and correlation tour allow rotation toward "interesting" projections of the data through projection pursuit (Cook et al. 1993). The link between the two programs allows the analyst to brush points, in either Arcview or XGobi, with a color/size/glyph and to see where the corresponding points are located in the other application. Thus, outliers in an XGobi plot can be brushed to see (in Arcview) where they were collected, or a spatial region in Arcview can be brushed to see (in XGobi) where the corresponding attribute measurements fall in the data space. Together, these tools provide a powerful and flexible environment for the graphical analysis of spatial data.

In addition to these basic capabilities, the link has been extended to include the display and analysis of SCDFs, spatially lagged scatter plots (Cressie 1993) and variogram clouds (Haslett et al. 1991, Bradley and Haslett 1992). In these cases, the data being passed from the GIS is processed before being displayed in XGobi. An explanation and examples of the SCDF link are given in the next section. The variogram-cloud link is used when exploring the spatial dependence in a data set and when looking for spatial outliers. In this option, the points displayed in XGobi represent all possible pairs of sampling locations. For each pair of locations, XGobi plots the square-root of the absolute difference between attribute values at the locations versus the Euclidean distance between the locations. In data sets exhibiting strong spatial dependence, the variance in the attribute differences will increase with increasing distance between locations. Locations that are near to one another, but with large attribute differences, might indicate a spatial outlier, even though the values at both locations may appear to be reasonable when examining the data set non spatially.

Figure 1 shows a variogram-cloud plot for precipitation sampling stations in which several potentially outlying points have been brushed. Because each point in the XGobi window corresponds to a pair of sampling locations, when the points in XGobi are brushed the Arcview window shows each pair of sampling locations connected by a line. This is also shown in Figure 1. Notice that all of the outlying points have a single sampling location in common. When the Arcview window is displayed with elevation contours, it is immedi ately obvious that the location in question is located on top of a mountain, which accounts for the large difference in precipitation.

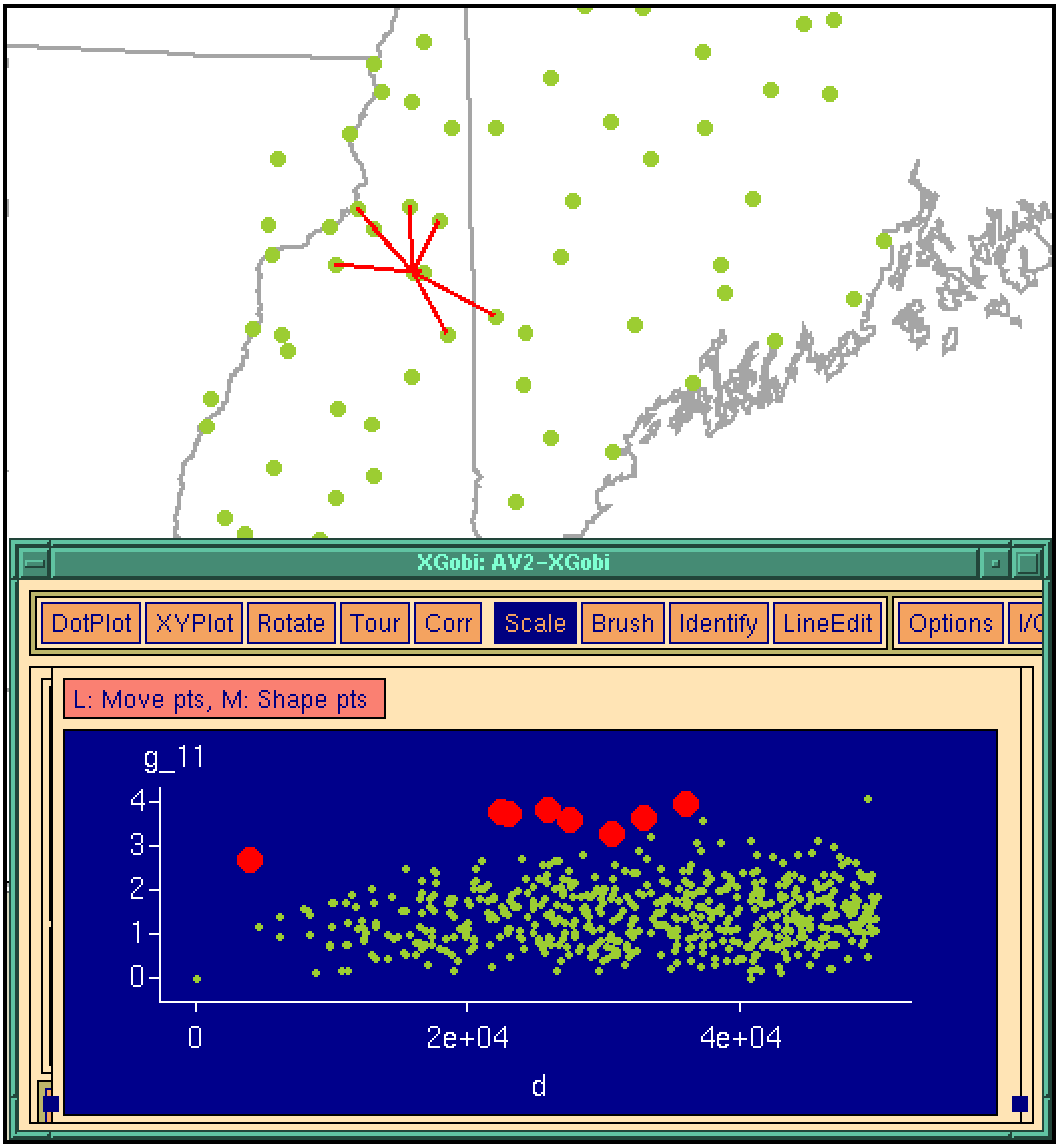


Figure 1. A variogram cloud plot (bottom of figure) with large values brushed. The map view (top of figure) indicates that all brushed points have a common sampling location.

PREDICTION OF THE SCDF FOR TREE CROWN HEALTH

In this section, the link described previously will be applied to the spatial prediction and visualization of the SCDF for the crown defoliation index (CDI) (Anderson et al. 1992), calculated from data collected in the northeastern United States. The CDI represents the nature of tree crown health as a response to stressors. In this analysis, the SCDF for the CDI process is predicted from data collected from a probability-based sample. Further more, we will use concomitant information, such as remotely sensed images, digital elevation models, and population densities, to improve the power of SCDF prediction for small areas. From the SCDF, it is possible to predict the area of forested land that falls in health classes (e.g., poor, marginal, good) as defined by the CDI. Using the link, SCDFs can be compared between regions or between the entire spatial domain and a subset of that domain.

Definition of the SCDF

Before we proceed, some background is necessary. Consider the spatial process

$$\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D\}, \quad (1)$$

where D represents the region of interest. Because we are interested in tree crown health, there is a scaling issue of when individual trees, after aggregation, begin to look like a forest. After suitable aggregation, one can represent the ecological index as a random field with continuous spatial index.

Because the field data were taken over a small study site, which we denote as Δ, we chose this as our standard area. Henceforth, we shall define Δ as the spatial support unit (SSU). Thus, at location ***s***, we have SSU Δ and $Z(\boldsymbol{s})$ defined over $\Delta(\boldsymbol{s})$.

The SCDF for this process is defined as follows:

$$F_{\infty}(z;B_0) \equiv \int_{B_0} I(Z(\boldsymbol{u}) \leq z)\, d\boldsymbol{u} / |B_0|;\ z \in \Re, \quad (2)$$

where $B_0 \in D$ is the forested portion of D, $|B_0|$ denotes the area of B_0, and $I(A)$ denotes the indicator function equal to one if A is true and equal to zero otherwise. Then the SCDF is the fraction of area in the region B_0 for which the value of the spatial process Z is less than a cutoff value z. This is depicted graphically in figure 2.

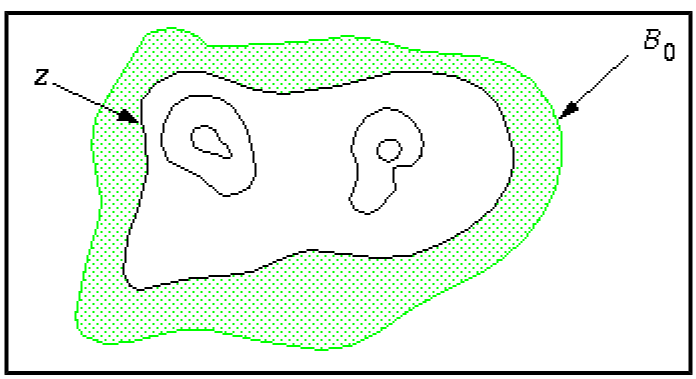


Figure 2. A graphical representation of the SCDF.

Because the information that we have is at a countable number of sampling locations and because we will use satellite data and other concomitant information to predict the SCDF, we shall tesselate the region B_0 into "tiles" made up of the image pixels. Let

$$B_0 = \bigcup_{i=1}^{N(B)} \{A(\boldsymbol{u}_i)\}, \tag{3}$$

where $A(\boldsymbol{u}_i)$ represents the image pixel defined at center point $\boldsymbol{u}_i$. There are $N(B_0)$ such pixels that make up B_0. For this analysis, then, we will use (3) and replace (2) with

$$F_\infty(z; B_0) \equiv \sum_{i=1}^{N(B)} I(Z(\boldsymbol{u}_i) \le z) / N(B_0), \tag{4}$$

where $Z(\boldsymbol{u}_i)$ refers to the crown index defined over $\Delta(\boldsymbol{u}_i)$ located at the point $\boldsymbol{u}_i$; i=1,..., $N(B_0)$.

Notice that we have effectively replaced the process $\{Z(\boldsymbol{s}): \boldsymbol{s} \in D\}$, with a discrete process

$$\{Z(\boldsymbol{s}_i): i = 1, \ldots, N(D)\}, \tag{5}$$

where $N(D)$ is the number of pixels that tesselate D in a manner analogous to (3). This discretization is essential for making progress but does introduce an approximation, the effect of which deserves further study.

Available to the researcher are data from the field,

$$\boldsymbol{Z} \equiv (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))', \tag{6}$$

obtained at sampling locations $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$. Given these data, a basic predictor of (4) is

$$\hat{F}_n(z; B_0) = \sum_{i=1}^{n} u(\boldsymbol{s}_i) I(Z(\boldsymbol{s}_i) \le z) / \sum_{i=1}^{n} u(\boldsymbol{s}_i), \tag{7}$$

where $\{u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_n)\}$ is a set of known weights, for example, inclusion probabilities in a sampling design. This is the form of the predictor that is used in this analysis.

Data

SCDF prediction will be examined for the crown defoliation index (CDI) of deciduous trees in the northeast United States. The data were collected as part of the Forest Health Monitoring program within the USEPA's EMAP. The CDI is the weighted average of two variables: crown dieback (*CDB*) and foliage transparency (*FTR*). The CDI for SSU $\Delta(\boldsymbol{s})$ is defined as:

$$Z(\boldsymbol{s}) \equiv \frac{\sum_{j=1}^{n(\boldsymbol{s})} DBH_j \bullet (CDB_j + FTR_j)/2}{\sum_{j=1}^{n(\boldsymbol{s})} DBH_j}; \; s \in B_0, \qquad (8)$$

where $n(\boldsymbol{s})$ is the number of trees at sampling location $\boldsymbol{s}$, DBH_i is the diameter at breast height of tree j; j=1,...,$n(\boldsymbol{s})$.

Crown dieback refers to the percentage of dead branches in the upper, sunlight-exposed parts of the tree crown. The assumption is that these branches have died from stressors in the environment other than lack of light. It is measured as a percentage in increments of 5 from 0% to 100%. Foliage transparency refers to the amount of light penetrating foliated branches. It ignores "holes" in the tree due to bare branches and is measured on the same scale as crown dieback.

The data were collected at sampling sites on the EMAP hexagonal sampling grid (White et al. 1992). The samples analyzied here were collected in the summer of 1992. In the study area, there are 66 sampling sites with deciduous trees.

The region under consideration is in the northeastern U.S. and includes portions of Maine, Massachussets, and New Hampshire. This region, which is shown in Figure 3, corresponds to the area of two Landsat satellite scenes.

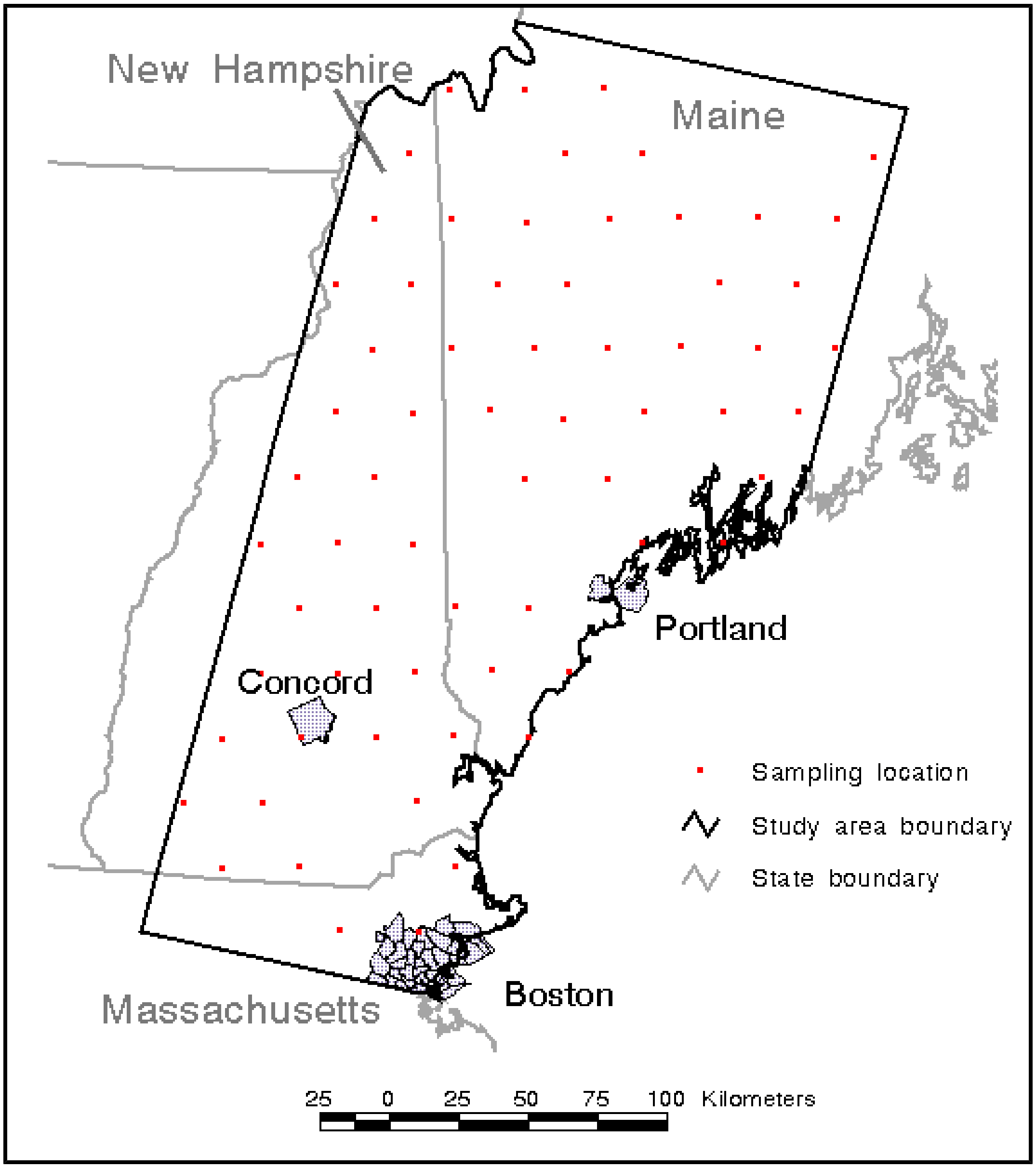


Figure 3. The study area.

Methodology

Our goal is to be able to predict the SCDF for small areas. In order to do this we shall exploit associations between sample data and data for which we have complete coverage, for example, remotely sensed data and digital elevation models. Observed associations will be used to predict values for the spatial process being studied at additional locations in the spatial domain. These points will then be used to predict the SCDF of the process for small areas.

The association between sampled data and the concomitant information is assumed to follow a simple linear model. Express the log of the CDI as the linear combination of concomitant variables plus a small scale stochastic term:

$$Y(\mathbf{s}) \equiv \log(Z(\mathbf{s})) = \mathbf{X}(\mathbf{s})\underset{\sim}{\beta} + \varepsilon(\mathbf{s}). \quad (9)$$

This model is fitted using weighted least squares regression, with the weights $w(\mathbf{s}_0)$ being equal to the sum of the DBH of trees at each location. The small-scale term is estimated from the residuals of the weighted regression model:

$$\hat{\varepsilon}(\mathbf{s}_i) = w(\mathbf{s}_i)^{1/2}\left(Y(\mathbf{s}_i) - \mathbf{X}(\mathbf{s}_i)\hat{\underset{\sim}{\beta}}_{wls}\right);\ i = 1, \ldots, n. \quad (10)$$

This term is assumed to be intrinsically stationary and can be predicted at any location, $\mathbf{s}_0$, in the spatial domain by:

$$\hat{Z}(\mathbf{s}_0) = \exp\left(\mathbf{X}(\mathbf{s}_0)\hat{\underset{\sim}{\beta}}_{wls} + w(\mathbf{s}_0)^{-1/2}(\hat{\varepsilon}(\mathbf{s}_0))\right), \quad (11)$$

where $\hat{\underset{\sim}{\beta}}_{wls}$ is the fitted regression coefficients from the large-scale model, $w(\mathbf{s}_0)$ is the weight for location $\mathbf{s}_0$, and $\hat{\varepsilon}(\mathbf{s}_0)$ is the predicted value for the small-scale term at location $\mathbf{s}_0$.

The Large-scale Model

The large-scale model is used to exploit associations between sample data and concomitant geographic information. This model was fitted using weighted least squares to express the log of the CDI for deciduous trees as a linear combination of regressor variables. The observations were weighted by the sum of the diameter at breast height for all deciduous trees at each location. The regressors that were considered include:

- *X and Y coordinates*: the coordinates of the sample locations (indicates a spatial trend)
- *precipitation*: the amount of precipitation in each of the four quarters prior to the sample date (predicted from NOAA precipitation values using optimal spatial prediction (i.e., a form of kriging))
- *greenness*: the greenness index of the tasselled cap transformation of Landsat remotely sensed imagery (Crist and Cicone 1984). This variable was calculated for the Landsat scene after a 3x3 average was applied to each pixel. The Landsat images used were acquired with the Landsat 5 sensor on June 12, 1993 (one year after the sample data were collected).
- *topography*: calculated from USGS 3-arc-second digital elevation models
  - *elevation*: the transformation, log(elevation), was used
  - *slope*: expressed in percent
  - *aspect*: the transformation, sin((1/2)*aspect), was used
- *population density*: the population density was derived from the 1990 U.S. census block groups; the transformation, log(population density), was used

Model selection

All possible models using the eleven regressor variables were fitted using weighted least squares. The final model was selected using four criteria:

1. low residual sum of squares;
2. high value of R squared;
3. significance of coefficients; and
4. low collinearity of regressor variables.

The colinearity of the regressor variables was evaluated using the condition index (Belsey et al. 1980). Any models with a condition index greater than 500 were not considered. Of the remaining models, the one with the lowest residual sum of squares and highest R-squared was evaluated based on the significance of coefficients. The goal is to find a model for which all coefficients are significantly different from zero at the 95% confidence level. This criteria was applied somewhat loosely, and the final model, which has a coefficient (the coefficient of the variable, sinaspect) that doesn't meet the criteria, is deemed acceptable. The largest condition index was 429.

The selected model is given below:

```
Residual Standard Error = 2.2517,  Multiple R-Square = 0.3395
N = 66,  F-statistic = 7.8396 on 4 and 61 df, p-value = 0

               coef     std.err  t.stat p.value
 Intercept -4.7035      1.9921 -2.3611  0.0214
         y  1.2783e-6  0.0000  3.4788  0.0009
 sinaspect  0.1551     0.0844  1.8381  0.0709
 greenness -0.0047     0.0022 -2.1381  0.0365
     p91q3  0.0534     0.0141  3.7868  0.0004
```

where `y` is the y coordinate, `p91q3` is the precipitation in the 3rd quarter of 1991, `greenness` is the Landsat greenness index, and `sinaspect` is the transformed aspect variable. The other variables were found to be unimportant according to the four criteria given above.

During the large-scale model fitting process, XGobi and the link between Arcview and XGobi were useful for several purposes. First, they helped in the exploratory spatial data analysis and the detection of the spatial outlier in the precipitation data set (see Figure 1). This data set was used to estimate the precipitation at each forest health sampling location. Second, through the use of the correlation tour, XGobi allowed us to check visually to see if there were associations between the explanatory and dependent variables and to check for collinearity among the explanatory vari ables. Finally, XGobi helped to assess visually regression diagnostics and outliers among the residuals.

Small-scale Model

The small-scale term of the linear model is estimated from the residuals of the weighted linear model; see (10). Variogram analysis on these residuals indicate that there is clear spatial structure. The variogram estimates, along with a fitted exponential variogram model, are shown in Figure 4.

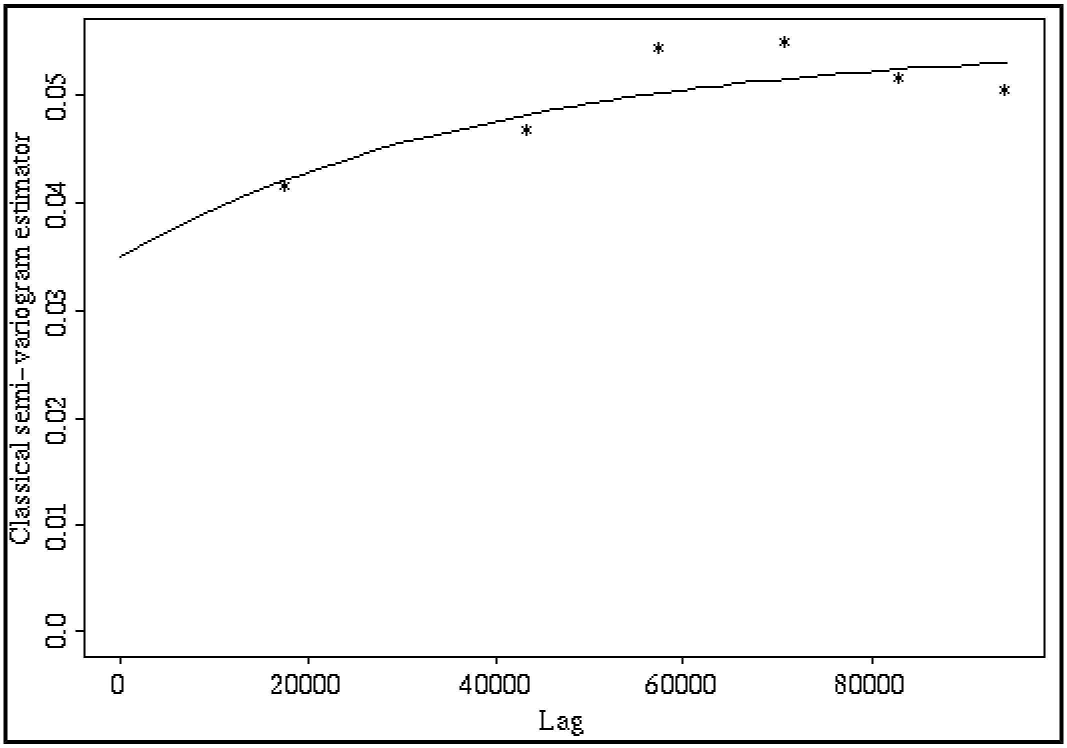


Figure 4. Variogram model fitted to the residuals of the fitted large-scale model.

When predicting the small-scale term for a location $\boldsymbol{s}_0$, constrained kriging (Cressie 1993) was used. Ordinary kriging involves the constraint

$$E(Z(\boldsymbol{s}_0)) = E(p(\boldsymbol{Z}, \boldsymbol{s}_0)), \tag{12}$$

where $p(\boldsymbol{Z}, \boldsymbol{s}_0)$ is the kriging predictor. It has been shown (written communication, Aldworth and Cressie) that ordinary kriging produces a process that is too smooth to be used for CDF prediction. Constrained kriging adds the additional constraint

$$var(Z(\boldsymbol{s}_0)) = var(p(\boldsymbol{Z}, \boldsymbol{s}_0)). \tag{13}$$

Together (12) and (13) match the first two moments of the predictor with the first two moments of the process. If we are to use the predicted values as if they were real data, as we do for SCDF pre diction, the additional constraint (13) becomes very important.

Determination of the Spatial Domain

Before SCDF prediction can be carried out, the spatial domain of interest, B_0, must be determined. In this case, B_0 is the portion of the study area that contains deciduous forests. For our analysis, we approximated this area by using the naturalized difference vegetation index (NDVI) and B_0 is defined as those areas for which the NDVI is greater than 0.5. This area is shown in Figure 5.

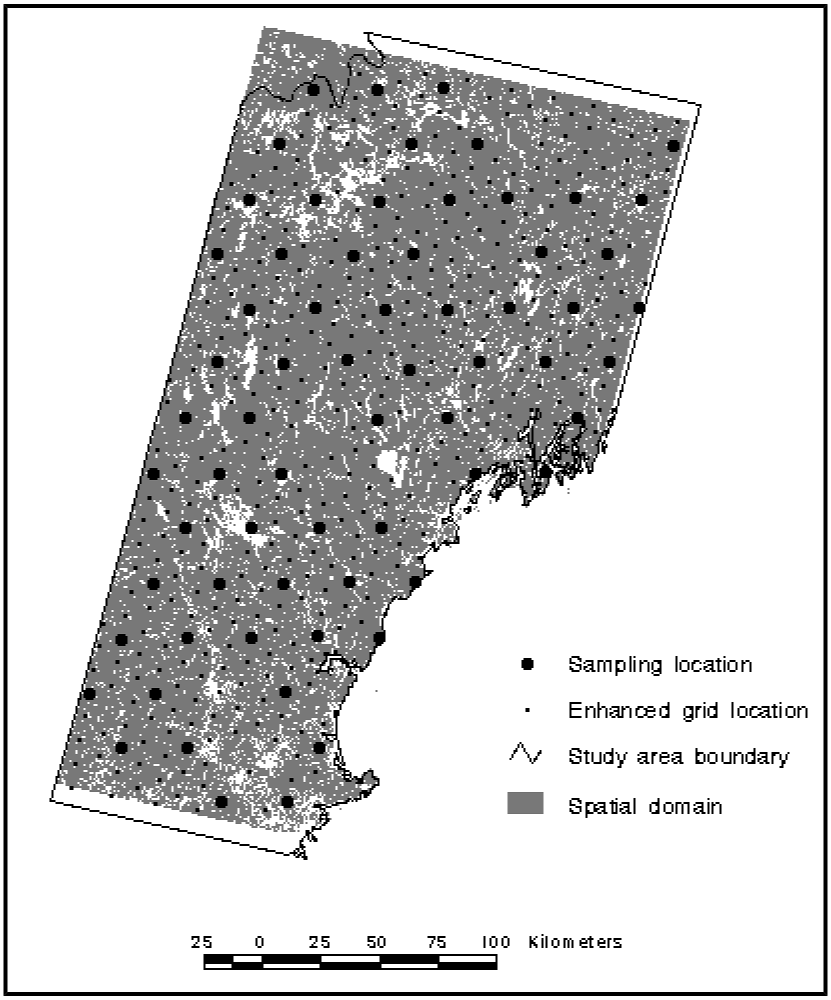


Figure 5. The approximated spatial domain of the CDI process in the study area. Also shown are enhanced grid locations.

Prediction of the Spatial Process at Additional Locations

After the preliminary work of model fitting and determination of the spatial domain has been completed, prediction and visualization of the SCDF can proceed. The first step is to predict the spatial process at additional points within the spatial domain. Points were added that correspond to a 7-factor enhancement of the original hexagonal sampling grid (White et al. 1992). This added 6 points for every point in the original grid (see Figure 5). Using (11) the CDI can be predicted for each new point that falls within the spatial domain. Because the weights used in (11), which are the sum of DBH for all trees at each location, are not known, they must be predicted. In this case, the weights were modeled as a function of the tasselled cap transformation (greenness) of the landsat image. The relationship between the two was determined by simple linear regression. The regression had an R-squared of approximately .25 and resulted in the following model to obtain the weights,

$$w(s_0) = greennes(s_0) \bullet 3.815.$$

Prediction and Visualization of the SCDF

The Arcview 2.1-XGobi SCDF link, introduced in the first section, can be used to predict, view and interactively query the SCDFs. This link provides several capabilities, including: (1) the defi nition of subregions of the spatial domain over which the SCDF will be calculated (up to 10 regions can be specified); and (2) linked brushing, in both directions, between the Arcview map window and the XGobi SCDF plot.

An example of an analysis using this link is shown in Figure 6. This figure shows the SCDFs cal culated for the CDI in two regions: that portion of the study area that falls in the state of Maine (the dashed polygon), and that portion that falls in New Hampshire and Massachusetts (the solid polygon). Figure 6b shows the predicted SCDFs for these regions; the SCDF on the left is for the New Hampshire/Massachusetts region and the SCDF on the right is for the Maine region. Figure 6b indicates that there is a difference in the CDI for the two regions.

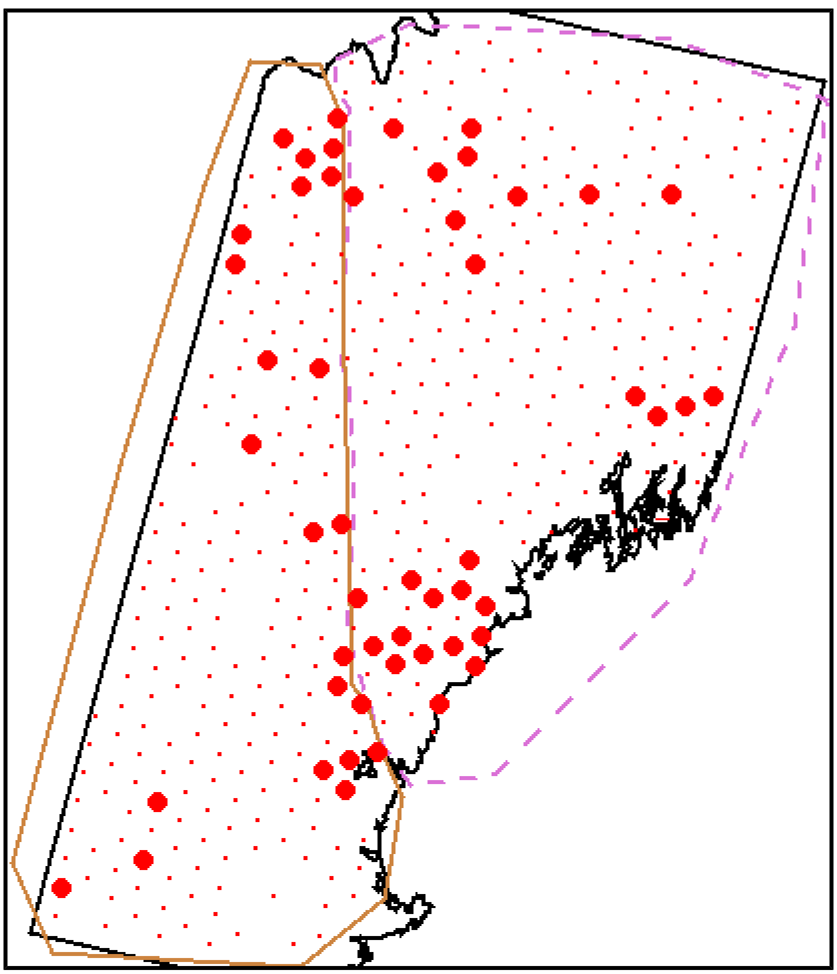


Figure 6a. Map of study area showing the regions that have been defined for CDF prediction.

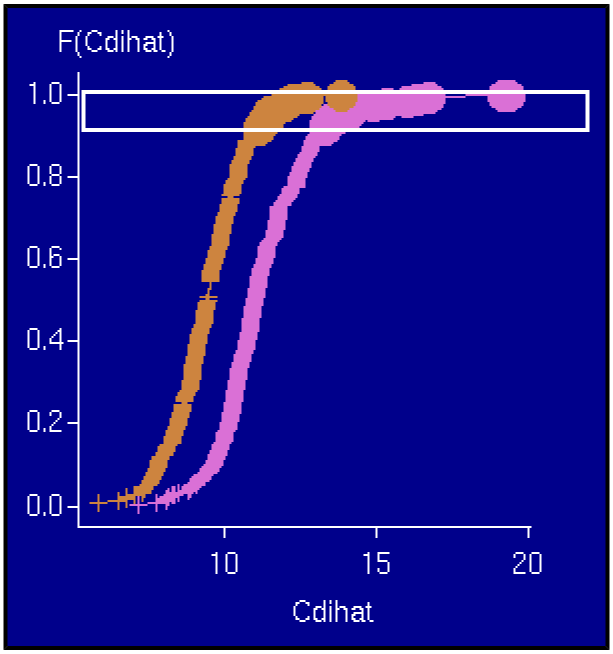


Figure 6b. Predicted CDFs for regions shown in Figure 6a. The horizontal, white box is being used to brush approximately the top 10% of the values in both CDFs. Brushed points are shown as large filled circles.

Figure 6 also gives an example of the brushing capabilities of the link. In this case, a horizontally shaped brush has been used to brush approximately the highest 10% of the values in both regions (Figure 6b). These points are shown in the map view as large filled circles, indicating the sam pling locations containing high values. By moving the brush up and down, various quantiles of the data can be explored. Alternatively, a vertically shaped brush could be used to brush specific ranges of values in the SCDF. This might be done, for example, if a priori cutoff values for the index were known that divide the resource into levels. In the current example, these cutoff values might correspond to health classes.

Acknowledgments

Research related to this artical was supported by an EPA EMAP grant under cooperative agree ment #CR822919. The article has not been subjected to the review of the EPA and thus does not necessarily reflect the view of the agency and no official endorsement should be inferred.

REFERENCES

Anderson, R.L., Burkman, W.G., Millers, I., and Hoffard, W.H. (1992) Visual crown rating model for upper canopy trees in the eastern United States. USDA Forest Service, Southeastern Region, Forest Pest Management. 15 pp.

Asimov, D. (1985) The grand tour: A tool for viewing multidimensional data. *SIAM Journal on Scientific and Statistical Computing*, 6(1): 128-143.

Belsey, D.A., Kuh, E., and Welsch, R.E. (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Colinearity.* Wiley, New York

Bradley, R. and Haslett J. (1992) Interactive graphics for the exploratory analysis of spatial data - the interactive variogram cloud. *2nd CODATA Conference on Geomathematics and Geostatistics. Sci. de la Terre, Ser. Inf., Nancy*, 1992, 31: 373-386.

Buja, A. and Asimov, D. (1986) Grand tour methods: an outline. *Computing Science and Statistics*, 17:63-67.

Cook, D., Buja, A., Calorera, J., and Hurley, C. (1995) Grand tour and projection pursuit. *Journal of Computational and Graphical Statistics*, 4(3), pp. 155-172.

Cressie, N. (1993) Aggregation in geostatistical problems. In *Geostatistics Froia '92*, Soares, A. ed, Kluwer, Dordrecht, Vol. 1, pp. 25-36.

Cressie, N. (1993) Statistics for Spatial Data. Wiley, New York.

Crist, E. P., and Cicone, R. C. (1984) A physically-based transformation of Thematic Mapper data-the TM tasseled cap. *IEEE Transactions on Geoscience and Remote Sensing*, 22(3): 256-263.

Haslett, et al. (1991) Dynamic graphics for exploring spatial data with application to locating global and local anomalies. *The American Statistician*, 45: 234-242.

Majure, J. J., Cook, D., Cressie, N., Kaiser, M., Lahiri, S., Symanzik, J. (1995) Spatial CDF Estimation and Visualization with Applications to Forest Health Monitoring, *Computing Science and Statistics*, Vol. 27, to appear.

Swayne, D. F., Cook, D., and Buja, A. (1991) XGobi: Interactive dynamic graphics in the X window systems with a link to S. In *ASA Proceedings of the Section on Statistical Graphics*, pp. 1-8, Alexandria, VA. American Statistical Association.

Symanzik, J., Majure, J. J., Cook, D. (1995) Dynamic graphics in a GIS: a bidirectional link between ArcView 2.0 and XGobi, *Computing Science and Statistics*, Vol. 27, to appear.

White, D., Kimerling, J., and Overton, S. (1992) Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartography and Geographic Information Systems*, 19(1): 5-21.

NOTES

1. Brushing refers to the ability to change the color/size/glyph of points in the graphics window. Linked brushing means that brushing conducted in any of the linked applications is immediately displayed in all others.

James J. Majure
Iowa State University GIS Support and Research Facility

218 Durham Center
Ames, IA 50011
http://www.public.iastate.edu/~majure
E-mail: majure@iastate.edu
Telephone: (515) 294-0526
FAX: (515) 294-1717

The changing face of agroecosystem characterization: Models and Spatial Data, the Basis for Robust Agroecosytem Characterization.

John D. Corbett, International Centre for Research in AgroForestry (ICRAF), Nairobi, Kenya.

Abstract:

One of the many challenges facing international agricultural research is the simple understanding of potential solutions to constrains and the cost of achieving those solutions. Identification of opportunities and constraints is the task of characterization. Modeling within a GIS offers a mechanism to integrate the many scales of data developed in and for agricultural research. Data access, including modeling results, expands to a "decision system" or decision tool which uses a mix of process models (where appropriate/possible) and simple rule-based systems to develop domains which link socioeconomic information (e.g. farmer survey results and census data) and biophysical data (growing season climate characteristics, soils, terrain).

An accurate spatial (and temporal) database enables the characterization of agroecosystems. This ability is vital in the developing world for efficient resource allocation in agricultural research. Agroecosystems are complex entities which span several levels or scales, with different processes dominating each scale. Therefore, a dynamic agroecosystem characterization requires both biophysical and socioeconomic data Characterization integrity is maintained by addressing particular objectives with specific information - information which may not aggregate up - or down - scale (e.g. the aggregate description of a complex of soils would not deliver a sensible "regional" characterization).

With spatially interpolated climate data, digital elevation models, and low resolution soils data in place, agroecosystem characterization commences with simple models used to differentiate growing season and off season characteristics. These "climate analog" models serve to describe the initial domain or target area for a range of priority setting steps, from sample design in diagnostic surveys and field trials to the identification of constraints and the number of people affected for institutional priority setting. Socioeconomic information - usually much more difficult to acquire - becomes critical in refining target domains as resource access, land tenure, cropping system, labor availability etc. dominate the land use system at higher resolutions.

Introduction:

The face of agroclimatic and agroecological characterization is changing as the reproducibility and flexibility of GIS tools loosen the former restrictions to data integration. For example, agroclimatic and agroecological zonation schemes are standard tools used to target agricultural research and to set research priorities because these zonation schemes offer a relevant and available information about target environments. Geographical Information Systems (GIS) have created a new environment to conduct similar research prioritization and targeting efforts. With GIS, we can greatly refine the methodologies by which "targets" are defined, while at the same time expanding the opportunities and potential for the accurate targeting of agricultural research efforts.

Scholes (1995) argues that climate is a first order "determinant" of ecosystem character, with edaphic factors second, followed by human intervention and other natural disturbances. This hierarchical determination of the constrains on ecosystem structure provides a useful model for characterizing agricultural environments. Beginning with the climate surfaces for monthly minimum and maximum temperature, mean precipitation, and mean evapotranspiration, the foundation is prepared for a scale-integrated and dynamic mechanism for supporting agricultural research efforts.

In the first section of this paper, growing season conditions at a macro or regional scale are evaluated. Two models were used to identify the growing season. This macro-scale analysis proved a useful means to quantitatively characterize growing season conditions. This analysis contributed to an ongoing regional planning effort for disaster mitigation over the greater horn of Africa region. In the second example, the macro-scale climatic surfaces are again used, but here they contribute to a more specific effort: first, the sample design of a farmer survey (along with air photo interpretation and census data) and, ultimately, due to the farmer responses, the agroclimatic interpretation (based on the climatic surfaces) was altered to better suit the maize specific objective. In the third section, the "data exploration tool" or DET is described from the viewpoint of the framework for further contributions to scale integrated research efforts.

These efforts support and are integrated into an environment seeking to provide accurate spatial information to researchers and to decision makers. The decision support framework allows access to the suite of data relevant to agricultural and natural resource management research, ranging from macro-scale climate information through to plot and household level detail.

Case: Greater Horn of Africa (GHA):

Accurate targeting of germplasm to appropriate areas greatly decreases the risk of crop failure, increases farmer confidence in governmental and NGO extension officers, and improves the return to an investment in agricultural research by providing a means to set research priorities.

Following a natural or political disaster, human suffering is compounded when farmers are forced to eat their reserve seed stock. In Rwanda, political unrest and social collapse preceded the food crises and seed for the following growing season was consumed. Replacement of this seed is vital. Providing farmers with seed adapted to their biophysical situation requires detailed climate information both to select appropriate varieties and to identify areas for seed bulking (the process of producing seed).

Corbett et al. (1995) used two models to identify the most suitable time of year for crop growth over the GHA. In the first model, we examined the ratio of precipitation to potential evapotranspiration (P/PE) for each month and sought consecutive months above a minimum ratio of 0.50. Because much of eastern Africa has a bimodal rainfall pattern, we then identified which of the two seasons had a higher P/PE ratio. For the second model, we identified the five consecutive months with the highest mean P/PE ratio, with no minimum ratio or rainfall onset "trigger" considered. For the remainder of this document, the first model (consecutive months of P/PE > 0.5) will be referred to as the "best" season model while the second will be referred to as the "optimum" season model (five consecutive months with highest mean P/PE). Though this nomenclature is not exact, the models serve different purposes depending on crop adaptation criteria - and, depending on the crop, one of the models will provide the more accurate growing season definition.

The best season for crop production is shown in Maps 1 and 2: Map 1 , the first month of the best season and Map 2 the length of this best season in months. Map 3 shows the first month of the optimum season. For both maps, we eliminated the area for which one or no months P/PE exceed 0.50 because these areas are beyond the rainfed margin for crop production (other conditions such as irrigation or water concentration would be necessary to produce grains and therefore these areas fall below the resolution of our database).

These two models are fundamentally different because of the agricultural significance in the identification of the first growing season month. In semiarid tropical Africa, the need to carefully select the first month with some trigger indicating rainfall onset is a priority. For well watered areas, production risk is more closely associated with nutrient competition. For crops which require fewer than about 100 days to mature (typically targeted for semiarid areas), the optimum model is probably inappropriate because there is no minimum criteria for the onset of the growing season and the first month of the optimum season model may be too dry for planting. For short season crops (from 60 to 120 days), the P/PE > 0.5 model would be more appropriate.

Both the optimal and best models have an inherent limitation for semiarid areas because monthly climatological data can provide little information on the distribution of precipitation over the season nor do climatological data provide information on inter-annual variability. For semiarid areas, climatological data can even be misleading. Because precipitation over the GHA is intimately linked to the location of the Inter Tropical Convergence Zone (ITCZ), some areas have the characteristic of extremely low precipitation if the ITCZ simply does not arrive and rather good rainfall for years when it does. For these areas, a low mean or climatological normal, which may seldom occur, can only indicate that water availability is limiting.

Our first order characterization criteria of P/PE > 0.50, though arbitrary, is a reasonable indicator of conditions which are approaching climatological suitability for crop production. In Map 4 we present information related to the uni- or bi- modal nature of the precipitation seasons. Western Kenya, normally considered as having "bi-modal" rains, is found to be uni-modal by our 0.50 criteria. The inter-modal 'dry' season in western Kenya had P/PE ratios which did not fall below 0.50. In addition, high mountain areas might have two seasons on the basis of low evaporative demand during the 'winter' months. This model did not consider temperature information though we inspected climate graphs (the 12 month sequence of precipitation, evapotranspiration, maximum temperature, minimum temperature) from a sample of sites to insure that our criteria selected the proper season for crop production.

In comparing Map 1 (first month of the P/PE > 0.50 season) and Map 3 , there are some striking differences. For much of the Ethiopian highlands, the optimum model (Map 3) would show May as the first month while the P/PE > 0.50 model (Map 1) shows March and April. We argue that the well-watered areas should use the optimum model not only because 5 months better approximates the number of months to mature high yielding longer maturing varieties, but also because a ratio of 0.50 is a nominal estimate of the onset of growing season conditions.

This database serves as a foundation for continued efforts to improve particularly the edaphic databases for the region. Use of the climatic surfaces enables targeting of potential areas for grain production by climatic adaptation and thus serves to make more efficient the efforts to collect and assimilate edaphic and human characteristics for those target areas.

Case: Maize Database Project for Kenya:

There are many agroclimatic and agroecological schemes published for Kenya. Two in particular are utilized by Kenyan institutions: Sombroek et al. (1982) and Jaetzold and Schmidts (1983). These generalized zonation schemes identify broad areas of similarity from which interpretations as to the specific adaptability of particular crops have been attempted. There is a critical need for a more specific targeting tool as these generalized zonation schemes fail because they are too general and because they represent the synthesis of many data but do not allow for the reworking or access to the data for a specific objective. For example, Hassan, Onyango, and Ruto et al. (1994) describe the large potential for intensification of maize production in the well watered highlands of Kenya (farmer yields are considerably lower than well managed on-farm "experimental" plots). From this information of apparent opportunity, their arose two questions: is KARI (Kenya Agricultural Research Institute) producing the right technologies for maize farmers? Are the technologies properly targeted? The agroclimatological / agroecological approach used by the Kenya Maize Data Base Project (MDBP) addressed these questions.

The MDBP used spatial analysis of climatic attributes to delineate maize adaptation zones (Corbett, 1994). The maize specific production zones were developed on the basis of environmental suitability and experts' definition of adaptability ranges for existing germplasm. Biological suitability alone is not sufficient for targeting interventions. Farmer specific, socio-economic, institutional, and policy factors interact with the physical situation to influence the process of technology transfer. The MDBP's agroclimatic zonation scheme provided the spatial frame for designing a geo-referenced survey of maize farmers' practices (Hassan et al., 1994c). Data from 1407 farmer interviews were then combined with climate and population data and maize cultivation intensity (based on air photo interpretation) to assess the suitability and adequacy of the existing maize production characterization and parallel resource allocation.

Monthly climate surfaces were constructed for Kenya using the thin-plate spline routines of Hutchinson (1991, 1995). Following a methodology adapted for Kenya but similar to Pollak and Corbett (1993), a cluster analysis was conducted utilizing the monthly precipitation and temperature surfaces. Because we knew the maize production density for Kenya (DRSRS, 1990) from aerial photography, only those cells coincident with the maize production region were included in the clustering.

The cluster interpretation and placement into maize specific adaptation zones reflects expert opinion on maize germplasm adaptation. The major differences between these agroclimate zones and the generalized zones (Sombroek et al 1982. and Jaetzold and Schmidts, 1983) begins with the crop specific nature of these zones and includes the ability to reproduce the zones from the baseline data (and thus improve their accuracy when more climate data become available when constructing the climatic surfaces), but also offers a flexible environment as different "rules" in the expert system can be easily evaluated and, finally, all of the data remain in a database connected to the resultant zones, readily accessible to the user. The zonation was then a "final" zonation scheme without the restrictions imposed by a cartographically based publication.

While biophysical suitability is important (the most broad "domain") for accurate targeting of agricultural research, at the farm level, the success or impact of a new technology is dependent on a suite of cultural, socio-economic, institutional, and policy level factors (Hassan et al, 1994a). Accordingly, data on farmers' practices and resources, farming system and the socio-economic conditions under which maize is produced were compiled from geo-referenced farmer and village level surveys. Georeferencing was necessary for resolving the false dichotomy between survey information and spatial data which are usually analyzed and presented separately.

The georeferenced survey data made it readily integrable into the spatial frame. An inadequacy in our initial agroclimatic zones was identified from the farmer surveys. Our initial zonation scheme followed the "traditional" broad categories of maize germplasm: lowland, mid-altitude, highland, and dry areas. KARI has released improved varieties for each of these four zones and our survey identified a divergence between farmer practices and existing zonal germplasn specifications. This divergence was highest in the mid-altitude zone, where more farmers planted late maturing maize hybrids, developed for the highlands, than the medium maturing varieties developed for the mid-altitude zone.

Inadequacy of these zones in characterizing maize germplasm requirements may be the reason behind the dissatisfaction of farmers with the varieties targeted for their respective environments. The spatial, flexible data structure allowed for iterative improvements in the zonation scheme. New boundary conditions were identified (from maize breeders) and a new zone was identified which separated the mid-altitude from the highland zone. Highland maize is quite particular as it copes well with cool temperatures but is highly susceptible to yield declines if the temperature is too warm. The mid-altitude zone was large, encompassing mean growing season temperatures from about 18 C to 25 C and was clearly not providing sufficient specificity to the maize breeders because farmers indicated that they preferred even local materials over much of the cooler part of the mid-altitude zone.

With the establishment of a "transitional" zone between mid-altitude and highland, and then a re-evaluation of the farmer responses, we found that 44 percent of Kenya's maize is produced under biophysical conditions for which there was no improved germplasm. Survey results show that, on average, maize reaches maturity within the range of 150-180 days in the transitional zone. At the time of this study, the KARI maize breeding program did not screen for materials requiring between 150-180 days. This occurred because the highland program in Kitale screens for materials of more than 180 days, while for the mid-altitude zone, both the Embu and Katumani station select for materials that mature in less than 150 days. This research effort, in concordance with the traditional maize adaptation regions, clearly denied efforts to improve materials for this transitional zone. These results also indicated a demand for a strategic research effort to develop appropriate maize varieties for the environment in Kenya which produces nearly half the maize in all of Kenya.

The results of the Kenya Maize Data Base Project confirm the appropriateness of an agroclimatological approach to planning and evaluation of agricultural research. Agroclimatological or agroecological approaches are usually more relevant than political, administrative units for defining natural resource management environments and that biological assessment is not sufficient, that socio-economic, cultural, institutional, and policy level factors greatly influence target groups.

An important result of this research was the establishment by KARI of a new breeding program at Kakamega to screen for the previously excluded 150-180 day maize materials targeted for the newly identified maize environment between the mid-altitude zone and the highland tropical zone (Hassan et al. 1994a). Map 5 shows the resultant maize specific zones and the location of the farmer interview sites (20 farmers were interviewed from each site).

The Data Exploration Tool (DET)

Institutionalizing the "memory" of the field work conducted in the third world is a vital step in achieving an assessment of the impact of agricultural research. Baseline studies form the foundation of future assessments, and without a means of institutionalizing the results of on-site analysis, there is no way, save through the same researcher, to support future impact assessments. Simple access to these data and their visualization improves the geographic knowledge of the user while facilitating the ability of a decision maker or new researcher to move toward a more thorough understanding of the complexities of the issues at hand.

The DET (Corbett and O'Brien, 1995) is a generalized GIS application tool which accesses gridded, environmental data, point data, and vector based information. Access to the gridded data enables the rapid construction of simple "empirical" models of conditions at a site or for a zone, or across a transect for the purposes, initially, of supplying scientists with responsibility for managing agricultural experiments, a mechanism for identifying the target domain for each experimental site. Initially focused on this first order determinant, climate, data in the DET include population density, several soils attributes (pH, depth, AWC etc.), and topography.

The identification of the similar areas (either from site characteristics derived from a specific location or from entered ranges for selected variables) is based on a choice of models. For example, a zone of similarity could be found using annual information, though for agriculture, this step serves more as a bridge for researchers and decision makers who might be more comfortable with annual data even though they know it is not representative of seasonal conditions. An example of a site characterization is shown in Map 6 .

Once a zone of similarity is found, the range of characteristics for that zone can be found using classic GIS tools. In other words, a zone identified as having a seasonal precipitation total of 500-800mm with maximum temperatures never exceeding 32 degrees C, could then be broken down into its component sub-zones using additional data on soils, minimum temperatures, and categories of precipitation. Or, for example, the range of human population density across this zone could be used to stratify a subsequent sample.

For the Future:

The use of spatial information in a digital environment allows for the integration of data from separate scales. Just as broad agroclimatic adaptations zones can be identified for the Greater Horn of Africa region, the same data, with a more specific objective, can be utilized in combination with farmer survey data, aerial photo interpretation, and census information to develop crop and germplasm specific target areas.

These data serve as the foundation for the vertical transfer of detailed scientific experiments to the decision maker or policy recommendations. For example, a model of soil organic carbon decay, developed from plot specific measurements, can be used to assess the vitality of a particular agricultural system spatially. If the system is found to be unsustainable under a specific set of biophysical characteristics, then that farming system under those conditions would require modification. This knowledge facilitates the adoption of improvements to the system through more accurate targeting of farms for which system adjustment is most suitable. It is only via simulation and robust spatial data that plot-scale knowledge can be assessed over space and the results contribute to the allocation and targeting of resources to resolve problems for which a resolution is clear. One of the challenges facing international agricultural research is the simple understanding of potential solutions to challenges and the cost of achieving those solutions.

The look of agroclimatic and agroecological characterization has evolved from generalized, cartographically restricted summaries to robust databases which integrate plot, household, community, and macro scale biophysical and socio-economic information. These databases become the heart of subsequent objective driven interpretations of the data offering flexibility and reproducibility while institutionalizing the results for future assessment.

Acknowledgments:

Many thanks to the Rockefeller Foundation for supporting the Kenya MDBP and the DET, UNEP for support toward the DET, and Jamie Wyant for his insightful comments and clarity of organizational comments for this paper.

A final thanks to the International Centre for Research in AgroForestry for their continuted

support.

Bibliography:

Corbett, J.D. 1994. Agroclimatic classification of maize production in Kenya: A multivariate cluster analysis approach. Chapter 3 The Kenya Maize Data Base Project, to be published by *CIMMYT (International Maize and Wheat Improvement Center*, Mexico), 1996.

Corbett, J.D., R.F. O'Brien, R.J. Kruska, and E.I. Muchugu, 1995. Agricultural environments of the Greater Horn of Africa - a database and map set for disaster mitigation. 9 pp text, 31 maps, plus database.

Corbett, J.D. and R.F. O'Brien, 1995. The Data Exploration Tool. An Arc/Info tool for accessing gridded data for environmental assessment and research prioritization objective. ICRAF (International Center for Research in AgroForestry), Draft, Nairobi, Kenya.

Department of Resource Surveys and Remote Sensing, 1990. *Long Rains Maize and Wheat Production Trends in Kenya*, 1985-1989. W.K. Ottichilo and R.K. Sinange, Technical report 137, September 1990.

Hassan, R., J. Lynum, J.D. Corbett 1994a. Geographic Information Systems Applications to Research Evaluation: The example of Maize from Kenya. Presented at the 4th Biennial Rockefeller Foundation Social Science Fellows Meeting, November, 1994, Addis Ababa, Ethiopia. To be published in proceedings.

Hassan, R. J.D. Corbett and K. Njorge, 1994b. Combining Geo-referenced survey data with agroclimate attributes to characterize maize production systems in Kenya. Chapter 4, The Kenya Maize Data Base Project, to be published by *CIMMYT (International Maize and Wheat Improvement Center*, Mexico), 1996.

Hassan, R., R. Onyango and J. Ruto, 1994c. Farmers' perception of priority problems in maize production and the relevance of maize research in Kenya. Draft, CIMMYT (International Maize and Wheat Improvement Center) / KARI (Kenya Agricultural Research Institute) Maize Data Base Project, KARI, Kenya.

Hutchinson, M.F., 1991. The application of thin plate smoothing splines to continent-wide data assimilation. In J.D. Jasper (ed.) 1991, *Data Assimilation Systems, BMRC Research Report* No. 27, Bureau of Meteorology, Melbourne 104-113.

Hutchinson, M.F., 1995. IJGIS.

Jaetzold, R. and H. Schmidt, 1983. Farm Management Handbook of Kenya. Ministry of Agriculture, Nairobi, Kenya.

Pollak, L.M. and J.D. Corbett, 1993. Using GIS datasets to classify Maize-growing regions in Mexico and Central America. Accepted for publication, *Agronomy Journal*, March 1993.

Scholes, R.J., 1995. The Environmental Determinants of African Terrestrial Biomes.

Presented to the MEDIAS International School, Nairobi, January 1995 for special issue, *Geo-Science*, 1996.

Sombroek, W.G., H.M. Braun, and B.J. Van der Pouw, 1982. Exploratory soil map and agroclimatic zones map of Kenya. Exploratory Soil Survey Report No. E1, Kenya Soil Survey, Nairobi, Kenya.

M.F.Hutchinson, H.A.Nix, J.P.McMahon and K.D.Ord

THE DEVELOPMENT OF A TOPOGRAPHIC AND CLIMATE DATABASE FOR AFRICA

The development of a gridded topographic and monthly mean climate data base for the entire African continent is described. The data will facilitate research planning, experimental design and technology transfer. The gridded climate data have been obtained by fitting thin plate smoothing spline functions of longitude, latitude and elevation to point values of monthly mean daily minimum temperature, daily maximum temperature and precipitation. The incorporation of a continuous spatially varying dependence on elevation is a critical factor in the accuracy of these surfaces. They interpolate monthly mean temperatures to within standard errors of about 0.5 degrees Celsius and monthly mean precipitation to within errors of about 10-30%. The interpolation of a digital elevation model at a resolution of 0.05 degrees latitude and longitude (approximately 5 km) has enabled the interrogation of the elevation dependent climate surfaces on continent-wide grids which have been released on CD. The work has involved a considerable effort in climate data collection, digitising of topographic data and extensive correction of data errors. The diagnostic procedures afforded by the interpolation procedures have played a crucial role in identifying errors in the primary topographic and climate data.

INTRODUCTION

The gridded African topographic and climate database has been constructed to address pressing problems associated with improving food production, managing pests and diseases and preserving biodiversity. The utility of this data base is based on the fact that climate plays a key role in determining constraints on population, livestock, cropping systems, and native flora and fauna. Topography in turn plays a key role in moderating the spatial distribution of climate. The thin plate smoothing spline method permits accurate topographically dependent interpolation of monthly mean climate from point data held in standard meteorological networks. Crucial to the accuracy of this procedure is the incorporation of a continuous, spatially varying dependence on elevation.

This paper demonstrates the application of the thin plate smoothing spline technique, as developed by Hutchinson (1991), to the African continent. This has involved splitting the continent into two overlapping tiles for the interpolation of temperature and fifteen overlapping tiles for the interpolation of precipitation, reflecting the greater spatial complexity of precipitation and the larger data network it requires. Estimated standard errors of the fitted surfaces reflect data network density and the spatial variability of the monthly mean climate.

The climate grids are calculated by applying the fitted smoothing spline surfaces to an underlying digital elevation model (DEM) with a spatial resolution of 0.05 degrees of longitude and latitude (approximately 5 km). The construction of this DEM is also demonstrated. This was obtained by applying the elevation gridding procedure ANUDEM (Hutchinson 1989) to elevation and streamline data digitised from 1:1M scale air navigation maps. The detection and correction of errors in both topographic and climate data were an important factor in achieving the accuracy of the final result. The DEM is a parallel development to the continent-wide DEM developed for Australia by Hutchinson and Dowling (1991).

The climate surfaces and grids can be used in a variety of ecological applications, including the prediction of spatial distribution of flora and fauna species using the BIOCLIM procedure developed by Nix (1986) and implemented in the ANUCLIM package (McMahon et al. 1995). The climate grids can also be used to classify the continent into ecological zones for the purposes of allocating land use and identifying priority areas for conservation of biodiversity.

TOPOGRAPHIC DATA

Topographic data were digitised from 39 1:1M scale air navigation charts covering the entire continent, augmented by miscellaneous maps at larger scales in areas where data on the air navigation charts were sparse. The digitised data consisted of:

1. All spot heights.
2. Selected points on elevation contours. All significant corners on contours, consistent with an eventual final grid resolution of approximately 1 minute of longitude and latitude, were selected. Thus points on contours were not sampled at spacings closer than 1 minute.
3. Selected stream lines. All streams which could be resolved on an eventual final grid resolution of 1 minute of longitude and latitude were selected. Only very minor streams on the 1:1M air navigation charts were omitted.
4. Coastlines of the continent and significant off-shore islands.

The point and line data were selected by manually tracing the data onto an overlying clear plastic film. The selected data were then digitised. The digitised data were rectified to longitude and latitude coordinates from the lambert conformal projection of the 1:1M air navigation charts using a program which calculated the mathematical inverse of the projection. The projections of the miscellaneous additional maps, which were mostly in transverse mercator projections, were similarly inverted. Positional errors of the rectified data were estimated to be no more than 1 km.

A total of 114,000 elevation points were digitised. This included 5,000 zero elevation points obtained from the digitised coastlines. The streamlines, when generalised to the final resolution of the DEM of 0.05 degrees latitude and longitude contained 220,000 points. In order to be used by the elevation gridding program the streamlines were digitised in the downslope direction.

CONSTRUCTION OF THE DIGITAL ELEVATION MODEL

The ANUDEM program developed by Hutchinson (1989) uses an efficient iterative finite difference procedure to interpolate a regular grid of elevations from point and contour elevation data and streamlines. An earlier version of the program has been implemented as the TOPOGRID tool in ARC/INFO. The procedure enforces drainage enforcement on the fitted DEM subject to elevation tolerances, in recognition of the fact that natural depressions are rare in most landscapes. The drainage enforcement sub-procedure adds significantly to the accuracy of the DEM, particularly when using limited data. Remaining depressions in the DEM were the main means of detecting errors in both elevation and streamline data. This provided an efficient way of detecting subtle errors in elevation data. Such errors may not show up as outliers in the usual statistical sense, but can be easily identified when they impede surface drainage.

The latest version of the ANUDEM program automatically smooths the natural discretisation error associated with the incorporation of elevation data onto a regular grid. This depends on the slope of the DEM, in a spatially varying manner, and the grid spacing. By monitoring the slopes of the DEM at the elevation data points, it is possible to optimise the resolution of the DEM to the information content of the source data (Hutchinson 1996). The natural discretisation error provides a direct estimate of the error of the DEM. The root mean square discretisation error of the fitted 0.05 degree resolution DEM, shown in Figure 1, was 35m. A spatially variable estimate of the standard error of the DEM, which depends on terrain roughness, ranges between about 20 and 150 metres. Hutchinson (1996) shows that the source data would support a DEM resolution as fine as 0.025 degrees. However the existing resolution creates a DEM of manageable size and is sufficient to support the spatial detail of the associated climate grids.

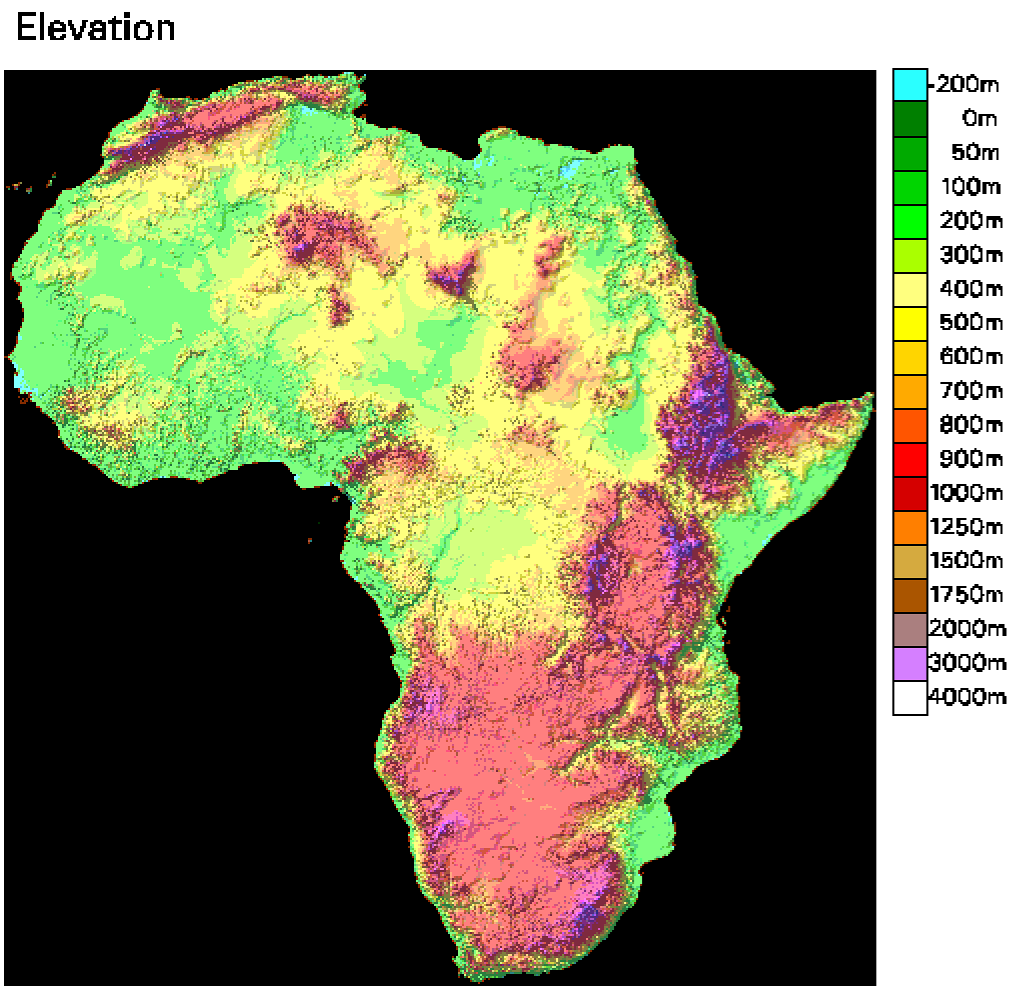


Figure 1. Digital Elevation Model of Africa.

A simple validation of the DEM was obtained by deriving the major streamlines from the DEM. These were determined by calculating the catchment area upstream from each grid cell and plotting in Figure 2 those cells which had upstream catchment areas exceeding 300 grid cells (approximately 7,500 km2). The streamlines shown in Figure 2 correspond closely with the known streamline network for Africa. Unconnected streamlines correspond with known major depressions in arid areas, including Lake Chad in the north and Lake Ngami in the south.

Figure 2. Major streamlines derived from the African DEM.

CLIMATE DATA

Monthly mean values of rainfall, daily minimum temperature and daily maximum temperature at a sufficient spatial density to support reliable spatial interpolation were compiled from a wide range of sources. In addition to data already obtained by CRES, monthly climate data were acquired from research agencies including CIMMYT, FAO, East Anglia Climate Research Unit, CSIRO Division of Forestry, Texas A&M University and from the national meteorological services of Djibouti, Gambia, Ghana, Kenya, Malawi, Morocco, Nambia, Rwanda, Seychelles, Sudan, Tanzania, Uganda and Zaire.

Data were collected over all available years of record to maximise spatial coverage, subject to the condition that rainfall averages were for at least five years or record. Most data were collected between about 1920 and 1980 for both temperature and rainfall, so the fitted climates grids can be interpreted as estimates of standard means for the period 1920 to 1980.

The number of accurately geocoded stations for which monthly mean climate data were obtained were 1504 stations for daily minimum temperature, 1499 stations for daily maximum temperature and 6051 stations for precipitation. The locations of the temperature data points are shown in Figure 3 which shows the two tiles in which the temperature surfaces were calculated. The locations of the precipitation data points are shown in Figure 4 which shows fifteen tiles. The data used in fitting the surfaces for each tile included a two degree wide overlap of each neighbouring tile to remove edge effects.

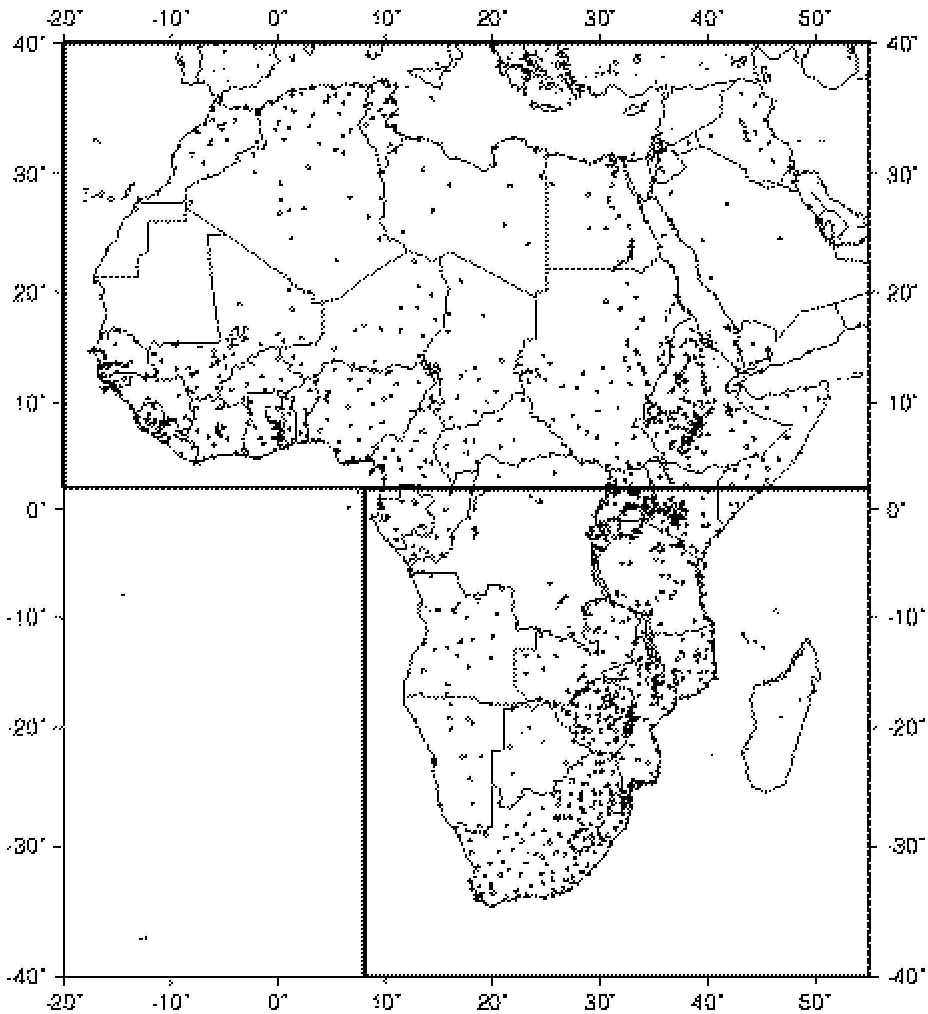


Figure 3. Temperature data locations in two tiles.

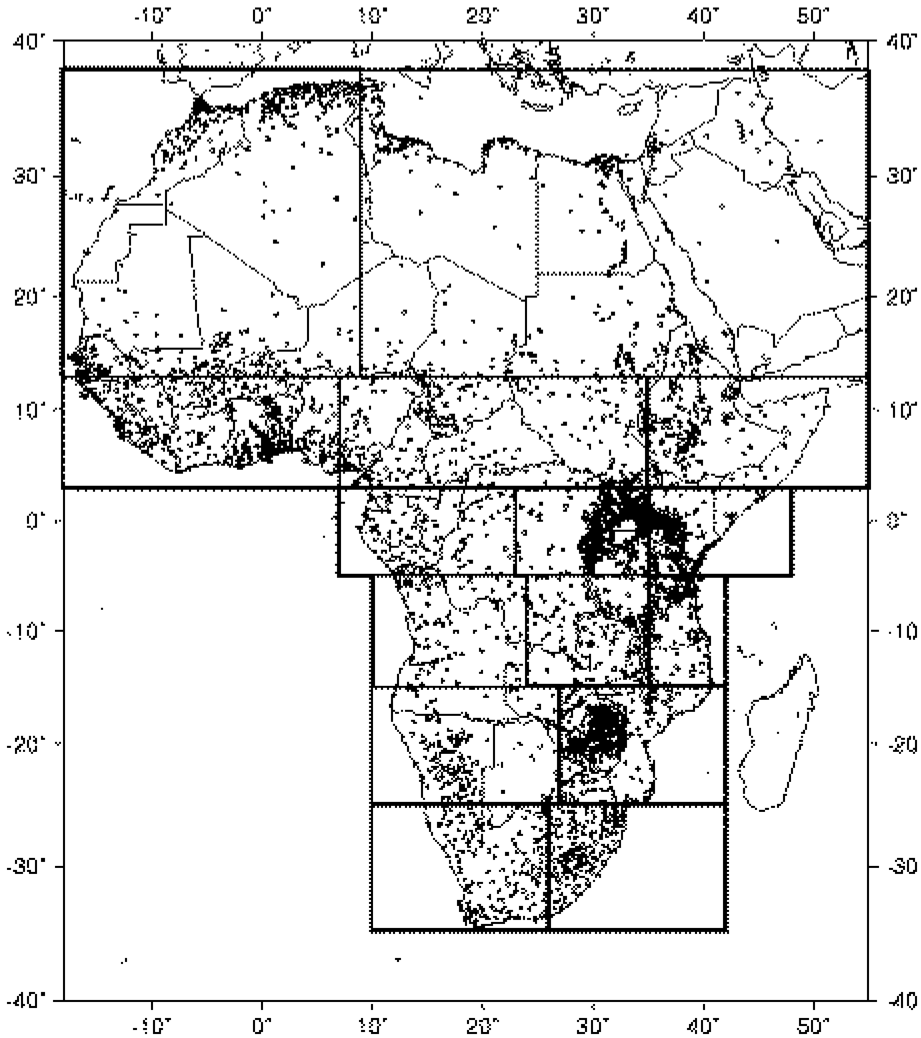


Figure 4. Precipitation data locations in fifteen tiles.

FITTING THE CLIMATE SURFACES

Procedures from the ANUSPLIN package (Hutchinson 1991) were used to fit the thin plate spline functions which were tri-variate functions of longitude and latitude in degrees and elevation in km. This relative scaling of the independent variables, which effectively makes the surfaces one hundred times more sensitive to elevation than to horizontal position, has been validated in precipitation analyses by Hutchinson (1995).

In fitting surfaces to noisy data, thin plate smoothing splines determine an optimal trade-off between goodness of fit and surface smoothness to minimise the generalised cross validation. The GCV is an estimate of the interpolation error obtained by removing each data point in turn and fitting a surface to the remaining data to see how well each omitted point can be predicted. The GCV is calculated implicitly and therefore computationally efficiently. More comprehensive discussion of statistical aspects of thin plate splines, and their relationship to kriging methods, may be found in Hutchinson (1993) and Hutchinson and Gessler (1994). A comparison of the thin plate smoothing splines with other methods of interpolating precipitation is being presented at this conference by Wilson and Sillman (1996).

The climate data were subjected to comprehensive error detection and correction. Errors in station positions and elevations were detected by examining points with the largest residuals from the fitted thin plate spline surfaces, as given in the ranked residual list output by the ANUSPLIN programs. This quickly identified errors. For a substantial number of stations, elevations were either too approximate or missing. Accurate geocoding of these stations was completed using topographic maps at scales finer than 1:1M where possible.

The procedures calculate various summary statistics and diagnostics which include the data mean and standard deviation, the estimated number of parameters (or signal) of the fitted surface, the GCV and the MSE, which is an estimate of the "true" predictive error which removes the data error component of the GCV. A ranked list of the largest data outliers is also produced to facilitate error detection and knot selection, as described below.

The signal is a useful diagnostic tool. The signal should not normally exceed about half the number of data points (Hutchinson and Gessler 1994). When it does the data network may be too sparse, the statistical model may be mis-specified or there may be gross data errors. When fitting surfaces to monthly data there should be a steady progression in the values of the signal from month to month. Departures from this progression may indicate systematic data errors in the aberrant months.

For most tiles, the number of data points exceeded 500 and approximate versions of the thin plate spline procedures were used. These depend on choosing a set of "knots" which equi-sample the three dimensional independent variable space. An initial set of 300-400 knots, about one quarter to one third of the number of data points, were initially chosen by the SELNOT program in the ANUSPLIN package. An additional 25 knots were chosen from data points at the top of the ranked residual list, after data errors were identified and corrected. The surface was then fitted with the expanded knot set and another 25 knots were chosen from the top of the ranked residual list. The knot set for the final fitted surfaces had 350-450 points.

TEMPERATURE SURFACES

The temperature surfaces were fitted for each tile using the SPLINB program. This program fitted twelve monthly temperature surfaces simultaneously, with each monthly data set weighted uniformly. This is the appropriate weighting for temperature means for which year to year variability is not a large factor in determining data errors. Errors are defined here as departures from the long term 1920-1980 mean.

Table 1. Summary statistics for the mid-season months for the maximum temperature surfaces for the northern tile. All statistics except the signal are in units of degrees Centigrade. There were 917 data points, including the data points in the two degree overlap with the southern tile. There were 450 knots.

			Data Std	Square	Square
Month	Signal	Data Mean	Deviation	Root GCV	Root MSE
January	331.0	26.94	6.67	1.23	0.59

April	365.5	30.55	6.06	1.28	0.63
July	395.1	30.31	6.00	1.47	0.75
October	336.5	29.90	4.66	1.27	0.55

The summary statistics for the maximum temperature surfaces for the northern tile are shown in Table 1. The signals are much less than the number of knots, indicating that a sufficient number of knots were chosen. There is also a smooth progression in the signals from month to month, with highest signals for the northern hemisphere summer. The square root of the GCV may be interpreted as a root mean square validation error. This is an overestimate of the true interpolation error because the data points used in calculating the GCV have errors. The square root of the MSE is an estimate of the true interpolation error after the effects of errors in the data points are removed. The sparse data network and uneven data quality has yielded true interpolation errors only slightly larger than the 0.5 degree Centigrade error typical of denser data networks (Hutchinson 1991). The estimated true interpolation errors for the southern tile, for which there is more uniform data coverage, ranged between 0.4 and 0.5 degrees Centigrade.

A plot of January mean daily maximum temperature for the entire continent is shown in Figure 5. The ESOCLIM procedure from the ANUCLIM package (McMahon et al 1995) was used to calculate the January maximum temperature grid using the DEM and the surfaces for each tile.

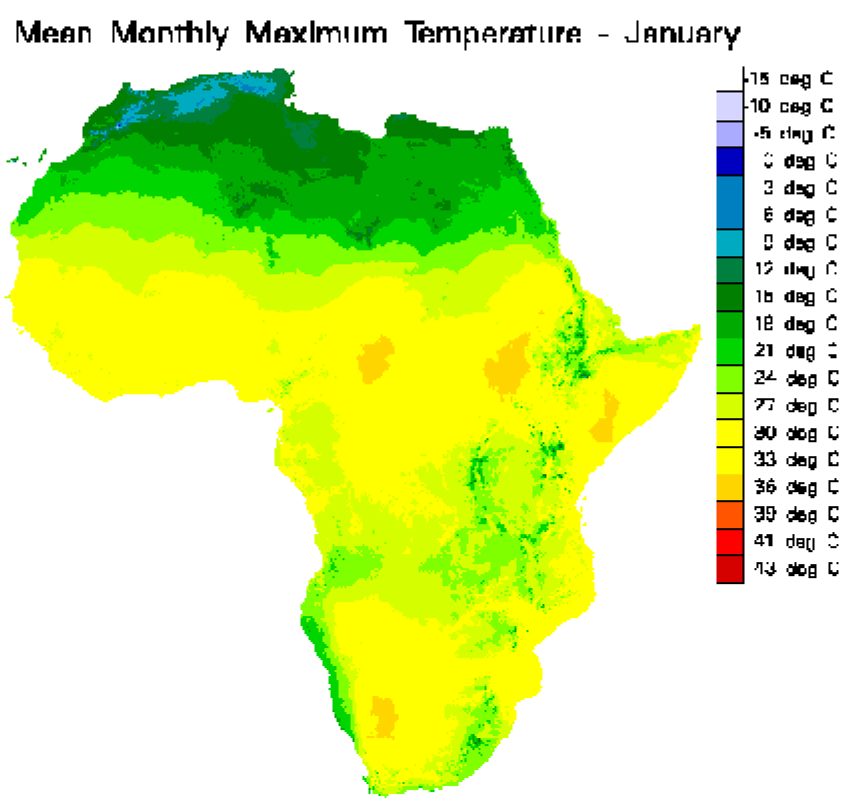


Figure 5. January mean daily maximum temperature.

The estimated true interpolation errors for the minimum temperature surfaces were slightly larger than the corresponding errors for the maximum temperature surfaces, reflecting more complex controls on minimum temperature, especially in winter where cold air drainage can locally invert the broad scale temperature environmental lapse rate. Nevertheless, the minimum temperature interpolation errors ranged between just 0.5 and 0.8 degrees Centigrade over both tiles.

PRECIPITATION SURFACES

The precipitation surfaces were fitted for each tile using the SPLINBB program. This program fitted twelve monthly surfaces simultaneously, but with each monthly mean data set weighted according to the approximate local error variance estimate

VAR/N

where VAR is the year to year variance of the monthly precipitation totals and N is the number of years of record. This weighting factor was introduced by Hutchinson and Bischof (1983) and discussed in Hutchinson (1995). It is appropriate for precipitation means for which year to year variability is a significant factor in assessing data errors. Since actual monthly data were unavailable for many stations, the variances were estimated by the formula

$$VAR = 1.77 * (MEAN + 1)^{1.649}$$

where MEAN is the monthly mean in millimetres. This formula was obtained by regression of the logarithm of monthly precipitation variances on the logarithm of the monthly precipitation means for Australia and holds reasonably well world-wide for the purpose of fitting monthly mean precipitation surfaces. The formula can be applied to both positive and zero monthly means.

Table 2. Summary statistics for the mid-season months for the precipitation surfaces for the tile with longitude limits 18 deg W to 9 deg E and latitude limits 3 deg N to 15 deg N (see Figure 4). All statistics except the signal are in units of millimetres. There were 640 data points, including the data points in the two degree overlaps with the neighbouring tiles. There were 350 knots.

			Data Std	Square	Square
Month	Signal	Data Mean	Deviation	Root GCV	Root MSE
January	231.1	11.61	16.90	7.0	3.3
April	168.0	82.84	67.04	22.1	9.7
July	202.2	229.89	165.10	54.5	25.4
October	190.5	147.72	104.09	35.2	16.1

The summary statistics for the precipitation surfaces for one data tile are shown in Table 2.

The signals are well below the number of knots, again indicating that sufficient knots were chosen. There is a reasonable progression in the signals from month to month despite the twenty-fold variation in data means from dry season in January to wet season in July. The estimated true interpolation errors, as given by the square roots of the MSE, range between 10% and 30% of the corresponding data means. Interpolation errors for remaining tiles are within similar limits. Largest percentage errors are for mid wet season in the Congo Basin, with high rainfall and minimal data coverage, and for dry seasons in desert and neighbouring regions, where percentage errors tend to overstate the significance of errors which are quite small in absolute terms. A plot of July mean precipitation is shown in Figure 6.

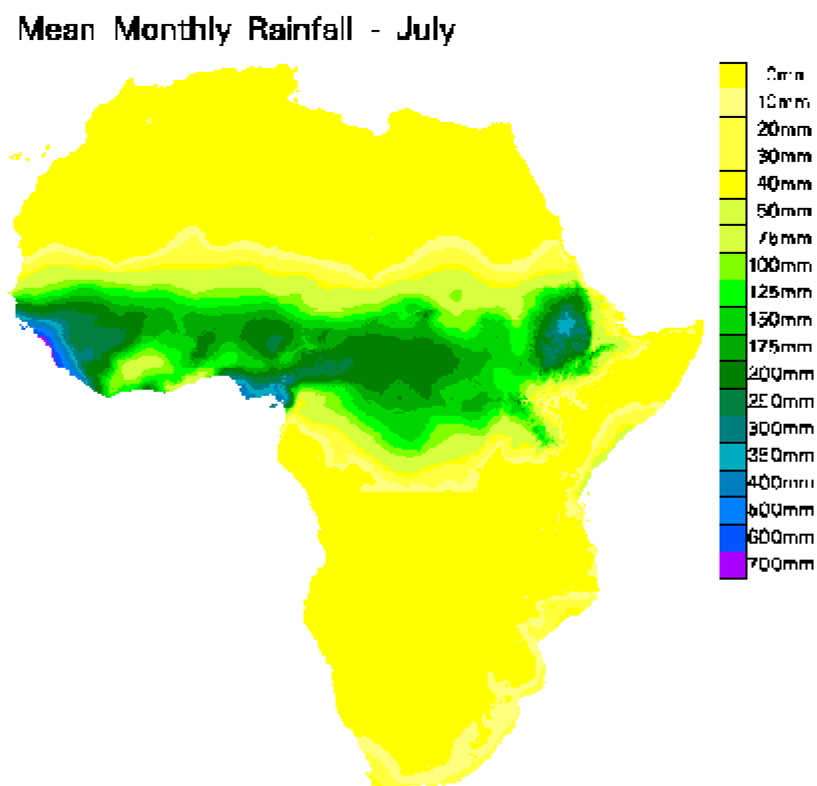


Figure 6. July mean precipitation.

DATA DISTRIBUTION ON COMPACT DISK

To facilitate distribution of this comprehensive data set, the DEM and the climate grids have been compiled on a compact disk (Hutchinson et al. 1995). The CD contains forty ASCII files containing gridded values of elevation (DEM) and annual and monthly mean values of daily minimum and daily maximum climate for the African continent at a spatial resolution of 0.05 degrees of longitude and latitude. Accompanying the ASCII grid files are forty colour images in device independent GIF format.

The ASCII grid files are in standard ARC/INFO ASCII GRID format. The CD also contains

DOS compatible programs for extracting data from the ASCII grid files in two different formats. The CD may be obtained from the Centre for Resource and Environmental Studies at the Australian National University.

ACKNOWLEDGMENTS

Funding for the project was provided by the Australian International Development and Assistance Bureau (now AUSAID) in response to requests by the International Laboratory for Research into Animal Diseases (ILRAD) in Nairobi, Kenya and the International Institute for Tropical Agriculture (IITA) in Ibadan, Nigeria. Digitising and checking of topographic and climate data formed the major portion of the work required to produce this data base. In this the authors gratefully acknowledge many willing contributors, including Fareeha Ibrahim, Helen Neave, Christy Sanders, Joseph Tabone and Karen Weinman. The authors acknowledge the generous assistance of Jennifer Kesteven in the preparation of the colour figures and the compact disk.

REFERENCES

Hutchinson,M.F. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* 106: 211-232.

Hutchinson,M.F. 1991. The application of thin plate splines to continent-wide data assimilation. In: J.D.Jasper (ed), *Data Assimilation Systems*. BMRC Res. Rep. No. 27, Bureau of Meteorology, Melbourne, 104-113.

Hutchinson, M.F. 1993. On thin plate splines and kriging. In: M.E.Tarter and M.D.Lock (eds), *Computing and Science in Statistics, Vol. 25*, Interface Foundation of North America, University of California, Berkeley, 55-62.

Hutchinson,M.F. 1995. Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographic Information Systems* 9: 385-403.

Hutchinson,M.F. 1996. A locally adaptive approach to the interpolation of digital elevation models. These proceedings.

Hutchinson,M.F. and Bischof,R.J. 1983. A new method for estimating the spatial distribution of mean seasonal and mean annual rainfall applied to the Hunter Valley, New South Wales. *Australian Meteorological Magazine* 31: 179-184.

Hutchinson,M.F. and Dowling,T.I. 1991. A continental hydrological assessment of a new grid-based digital elevation model of Australia. *Hydrological Processes* 5: 45-58.

Hutchinson,M.F. and Gessler,P.E. 1994. Splines more than just a smooth interpolator. *Geoderma* 62: 45-67.

Hutchinson,M.F., Nix,H.A., McMahon,J.P. and Ord,K.D. 1995. Africa - A topographic and climatic database, Version 1.0. Centre for Resource and Environmental Studies, Australian

National University, Canberra.

McMahon,J.P., Hutchinson,M.F., Nix,H.A. and Ord,K.D. 1995. ANUCLIM User's Guide. Draft Report, Centre for Resource and Environmental Studies, Australian National University, Canberra.

Nix,H.A. 1986. A biogeographic analysis of Australian elapid snakes. In R.Longmore (ed), *Atlas of Elapid Snakes of Australia*, Australian Flora and Fauna Series, No.7, Australian Government, Canberra, pp 4-15.

Wilson,J.P. and Sillman,S.T. 1996. Comparison of ANUSPLIN, MT-CLIM-3D and PRISM precipitation estimates. These proceedings.

AUTHOR INFORMATION

M.F.Hutchinson, H.A.Nix, J.P.McMahon and K.D.Ord
Centre for Resource and Environmental Studies
Australian National University
Canberra ACT 0200
Australia
email: hutch@cres.anu.edu.au

Trevor H. Booth

Predicting Plant Growth:

Where will it grow? How well will it grow?

Three recent projects have developed and enhanced generic methods to evaluate the suitability of particular plants for general regions and specific sites. All three projects have concentrated on forestry in developing countries, but the methods are suitable for other regions and for plants other than trees. Two of the projects have emphasized the development and application of cheap easy-to-use PC-based programs, whilst the third has shown how a simple PC-based model can also be applied as part of a multi-million dollar land evaluation study using the ARC/INFO GIS on Sun workstations. Four methods used in the projects are outlined here, including the development of climatic interpolation relationships, the development of climatic mapping programs, the application of the Plantgro simulation model and the development of simulation mapping programs. Though more work is needed to validate and calibrate these programs for the hundreds of tree species important in tropical and sub-tropical areas it is concluded that they already provide a useful basis for selecting species and provenances for planting in different regions.

Introduction

The world's rapidly rising population requires most countries to make the best possible use of their land resources for agriculture, horticulture, forestry and conservation. Being able to predict where and how well particular plants are likely to grow in different regions is vital for land use planning. Linking GIS and models can help to answer these questions, but decision makers and researchers in developing countries have limited access to these technologies. This paper outlines on-going efforts to provide improved access to databases, mapping programs and simple simulation models to assist land evaluation. The work described relates mainly to forestry projects in developing countries, including work for the Australian Centre for International Agricultural Research (ACIAR) project 9127 "Predicting Tree Growth for General Regions and Specific Sites in China, Thailand and Australia" and the Australian Agency for International Development (AusAID) project on "Improving Tree Productivity in Southeast Asia". However, the methods could also be applied in other areas and with plants other than trees.

Four methods are outlined. First, the development of climatic interpolation relationships which allow mean climatic conditions to be estimated reliably for any location within a particular country. Second, the development of climatic mapping programs which use interpolated data to indicate where particular trees can be grown. Third, the use of the Plantgro model, which uses detailed soil and climate information to estimate how well a particular tree is likely to grow on a specific site. Fourth, the use of simulation mapping programs, which use a simplified version of the Plantgro model, to predict how well a particular tree might grow at thousands of locations across a country or continent. The purpose of this paper is to provide a brief introduction to the methods. More details are available in the proceedings of an international workshop entitled "Matching Trees and Sites" (Booth 1996a).

Climatic interpolation

Climate has an important influence on plant growth. It is particularly useful as a means to predict where particular plants will grow, as mean climatic conditions can now be reliably estimated for most locations around the world. For example, Dr Michael Hutchinson (Centre for Resource and Environmental Studies, Australian National University) has developed a package known as ANUSPLIN, which uses Laplacian smoothing splines to interpolate spatially between data recorded at meteorological stations (Hutchinson 1989, 1992). As part of ACIAR project 9127 mean monthly data were collated from meteorological stations in a single area including China, Thailand, Vietnam, Laos, Cambodia and Peninsula Malaysia (Zuo et al. 1996). Monthly mean data for 1222, 1117, 3832, 898 and 903 stations were collated for maximum temperature, minimum temperature, precipitation, solar radiation and evaporation respectively. Interpolation relationships were developed relating these monthly mean values to latitude, longitude and elevation.

Latitude and longitude are easily estimated for any location, but elevation must also be provided to interrogate the interpolated climatic relationships. A digital elevation model was prepared by Zuo et al. (1996) and monthly mean values for all five climatic factors were estimated for a 1/20th of a degree grid (5km approx) of approximately 400 000 points across China and mainland Southeast Asia. Example colour maps showing mean annual values of these five factors across China and mainland South East Asia are included in the ACIAR proceedings (Booth 1996a). Climatic interpolation analyses for Indonesia (Jovanovic and Booth 1996a) and the Philippines (Jovanovic and Booth 1996b), which were developed as part of work for the AusAID, are also described in the ACIAR proceedings. Details of the interpolation methods used, sample outputs and references to studies in other areas are provided by Hutchinson et al. as well as Kesteven and Hutchinson (both papers in these proceedings).

Climatic mapping programs

Very large climatic databases are impressive, but they are of little help for decision-making in developing countries if potential users cannot access them. In countries where average incomes may be no more than a couple of hundred dollars a year, few individuals or institutions have access to sophisticated GIS programs or powerful computers. Even if low-cost PC-based GIS programs, such as IDRISI (Eastman 1993), are available they may seem complex to many first time users. Climatic mapping programs were developed to provide very easy access to interpolated climatic information (Booth 1996b). Simple descriptions of ranges of climatic conditions are entered and the climatic mapping programs show which areas, if any, satisfy those sets of conditions. As climatic mapping programs were mainly developed for forestry studies the following set of six climatic factors were used initially:

a) mean annual precipitation (mm)

b) rainfall seasonality (uniform/bimodal, summer, winter)

c) dry season length (months)

d) mean maximum temperature of the hottest month (oC)

e) mean minimum temperature of the coldest month (oC)

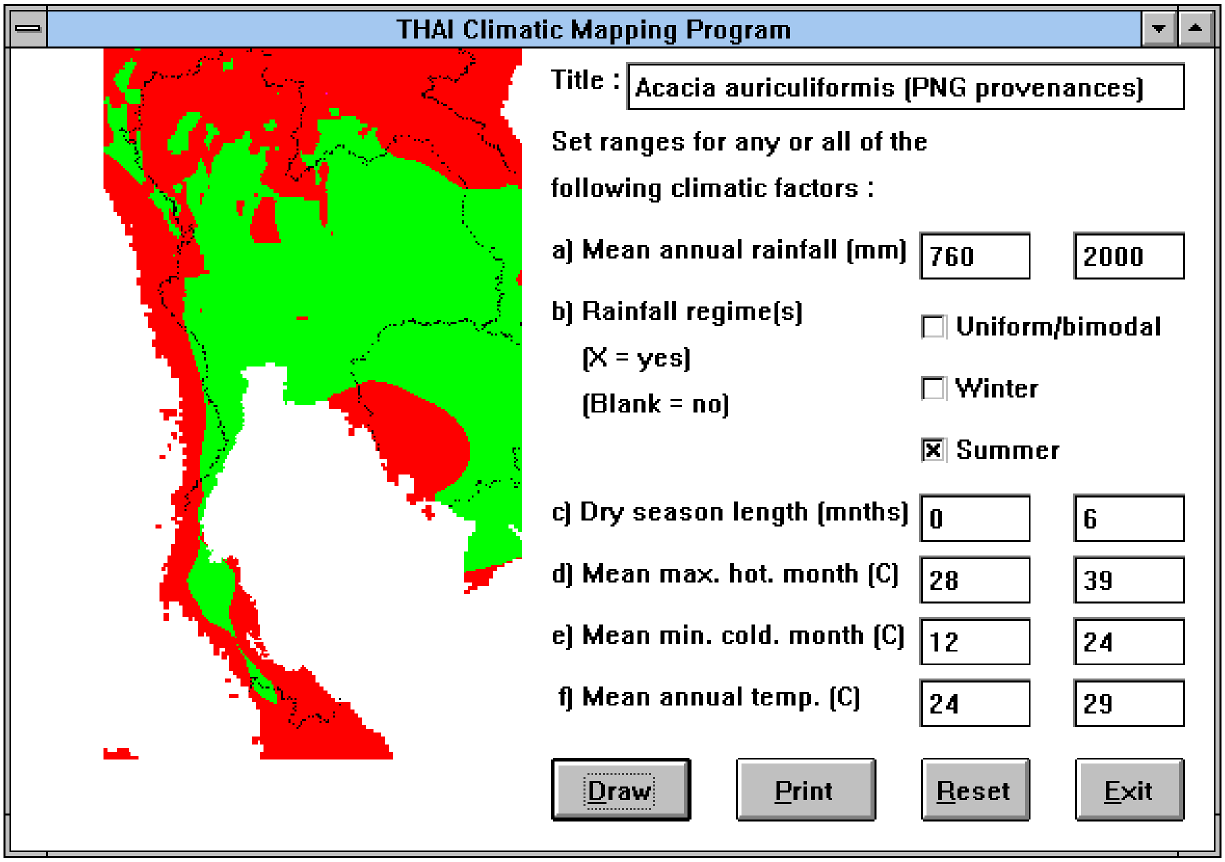
f) mean annual temperature (oC)

These factors had proved to be useful as part of an expert system developed to assist selection of tree species for tropical and sub-tropical plantations (Webb et al. 1980). In the second edition of their system, Webb et al. (1984) described the climatic requirements of 175 species in terms of ranges of these six climatic factors. Any one of these descriptions can be input into a climatic mapping program. Areas which satisfy the requirements are shown on the microcomputer screen in green, whilst locations which do not satisfy the description are shown in red. The user can move a marker over any location and check the detailed climatic conditions at a particular location.

Descriptions of the requirements of particular trees (or other plants and also animals) can be developed from bioclimatic analyses of natural distributions. For example, the BIOCLIM program accepts geocoded information describing natural distributions and uses climatic interpolation relationships to determine the range of climatic conditions where a particular plant or animal is found (Booth 1985, Nix 1986, Busby 1991). This type of analysis can provide a first indication of a particular tree's climatic requirements. However, many trees can grow successfully in conditions which are somewhat different from those they experience within their natural range. Information from trials and large scale plantations outside the natural range can be used to improve descriptions of requirements.

Climatic mapping programs, like more sophisticated GIS packages, allow users to visualise the implications of particular descriptions. For example, the Webb et al. (1980, 1984) description of the requirements of *Pinus radiata* implied that New Zealand was climatically unsuitable when it was plotted using a climatic mapping program for the whole world. As there are over 1.2 million hectares of *Pinus radiata* plantations in New Zealand something was obviously wrong with the description. Using the program's moveable marker it was easy to check conditions at sites in New Zealand and correct the errors in the description (Booth 1990). Working with individuals who have experience with growing trees in particular regions, climatic mapping programs can be used to quickly check and improve existing descriptions of trees' requirements (e.g. Booth and Pryor 1991).

As part of ACIAR Project 9127 climatic mapping programs were prepared at the CSIRO Division of Forestry for China (100 000 grid points), Thailand (40 000 grid points), South East Asia (10 000 grid points) and Latin America (66 000 grid points - interpolated climatic data kindly supplied by Dr Peter Jones, CIAT, Colombia). Most of the programs have been developed for the MS-DOS environment, which is the most common operating system on PCs used in the developing world. However, a version of the THAI program has recently also been developed for the Windows environment. Windows allows multitasking, which makes it easy to compare maps produced by different descriptions on the computer's screen, as well as providing built-in support for hundreds of different printers. Figure 1 shows the areas of Thailand which are climatically suitable for provenances of *Acacia auriculiformis* from Papua New Guinea.



As part of an AusAID-supported project climatic mapping programs were developed for Indonesia (26 000 grid points), Vietnam (16 000 grid points) and the Philippines (13 000 grid points). Papers describing the development of these programs are all included in the ACIAR proceedings. Whilst a climatic mapping program satisfies a simple definition of a GIS as "a computer system capable of holding and using data describing places on the earth's surface" (ESRI 1992) many people would consider them too simple to be called a GIS. Whether they are a GIS or not they certainly provide a useful and appropriate means of delivering environmental information to users in developing countries. Users who have both the equipment and skills necessary to operate more elaborate systems find it easy to incorporate the interpolated climatic databases used by climatic mapping programs into GISs.

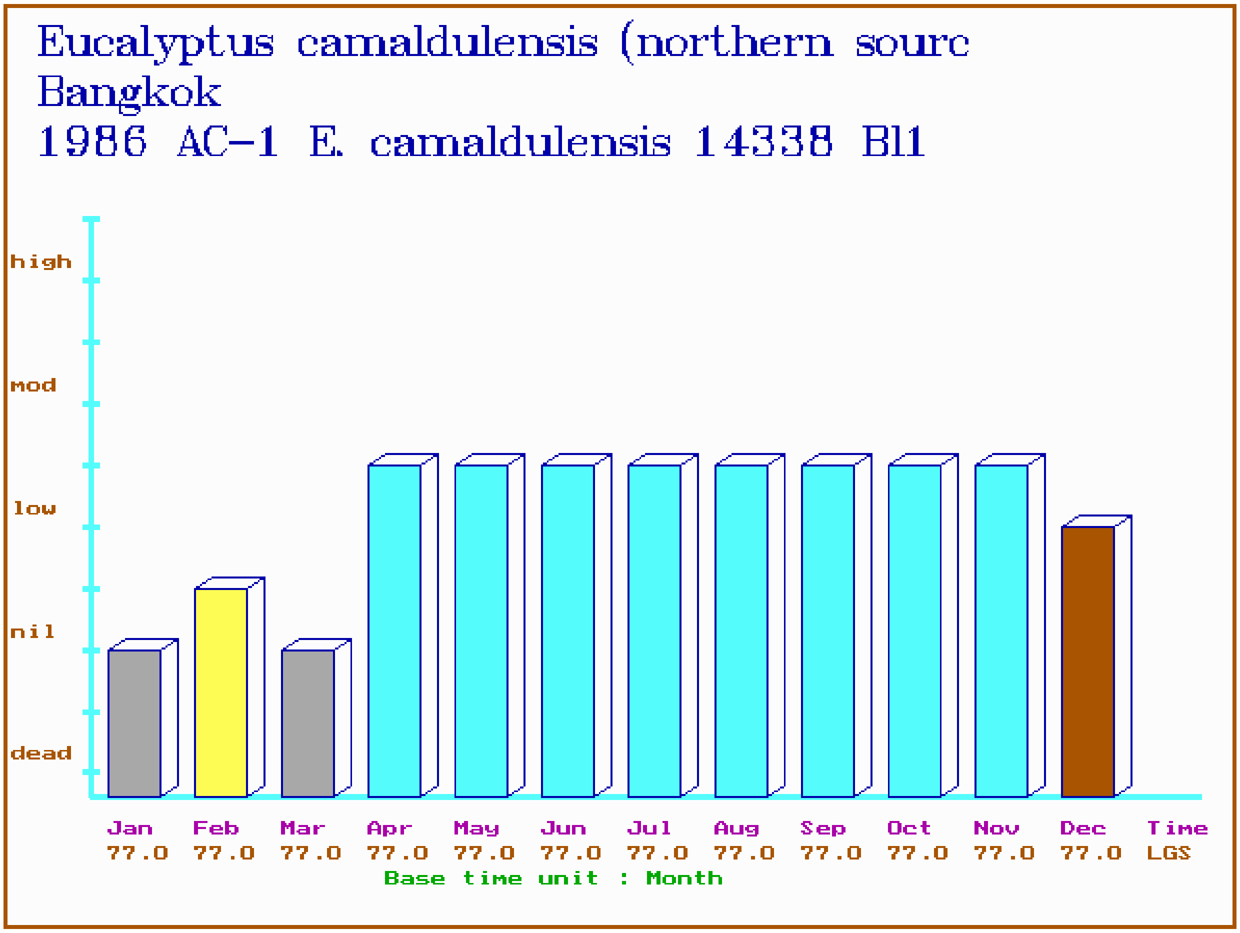
Plantgro

Being able to identify where particular trees (or other plants) will grow is useful, but many people need to know how well they will grow on particular sites. Generally they do not require highly precise predictions of yield, but they do need to know whether growth will be good, fair, poor or useless. Detailed process-based models are available for the dozen or so major crop plants, such as wheat and rice, which dominate world agricultural production (e.g. Godwin et al. 1989, Singh et al. 1993). These simulate complex processes such as light interception, photosynthesis and translocation and in many cases provide quite reliable estimates of yield. A few process-based models have been developed for trees (e.g. McMurtrie et al. 1989), but there is no prospect of such detailed models being developed for the hundreds of tree species which are important in forestry around the world. Dr Clive Hackett faced a similar problem in 1984 when he was asked to take part in a study of village-based subsistence agriculture and small-holder cash cropping in Papua New Guinea (Hackett 1988). There were numerous plant species involved and relatively little was known about their environmental requirements.

Hackett devised a new and simple method for providing coarse predictions of the growth of lesser-known plants. To assess the suitability of particular climatic or soil factors he used 'notional relationships', which are simply two-dimensional graphs made up of linear segments indicating conditions which are most suitable for growth and those which are less suitable. These are used along with more complex calculations of the effects of light, temperature and moisture. To combine the effects of all factors he used Liebig's Law of the Minimum (Liebig 1885), which was originally devised to describe the effects of available plant nutrients on plant performance. In simple terms this states that the most limiting factor determines plant performance (i.e. favourable levels of other factors do not compensate for the unfavourable level of the limiting factor). Overall conditions were evaluated according to limitation ratings on a 0-9 scale where 0 indicates ideal conditions (i.e. no limitations) and 9 indicates the greatest possible limitations.

The PC-based Plantgro program (Hackett 1991) evaluates 11 soil factors including phosphorus, potassium, nitrogen, slope and drainage. Climatic data used include maximum temperature, minimum temperature, precipitation, evaporation and solar radiation. Monthly mean data are usually used for trees, but the program can also evaluate ten-day or weekly data. The program evaluates the effects of temperature on development as well as carrying out simple water balance calculations. To run the program a plant file, soil file and climate file are required. The program provides summary predictions of likely growth patterns as well as detailed evaluations of limitations due to light, temperature, moisture and important soil factors.

Figure 2 shows an example of the summary output produced by Plantgro for a northern provenance of *Eucalyptus camaldulensis* growing at a trial site near Bangkok in Thailand. Growth is steady for much of the year, but is limited in the period from December to March. Inspection of Plantgro's detailed output would indicate that the growth limitations in the December to March period are mainly due to moisture stress because of the dry period, whilst growth during the rest of the year is limited at this particular site by soil depth. The '77.0's at the foot of the figure indicate that a perennial plant is being evaluated.



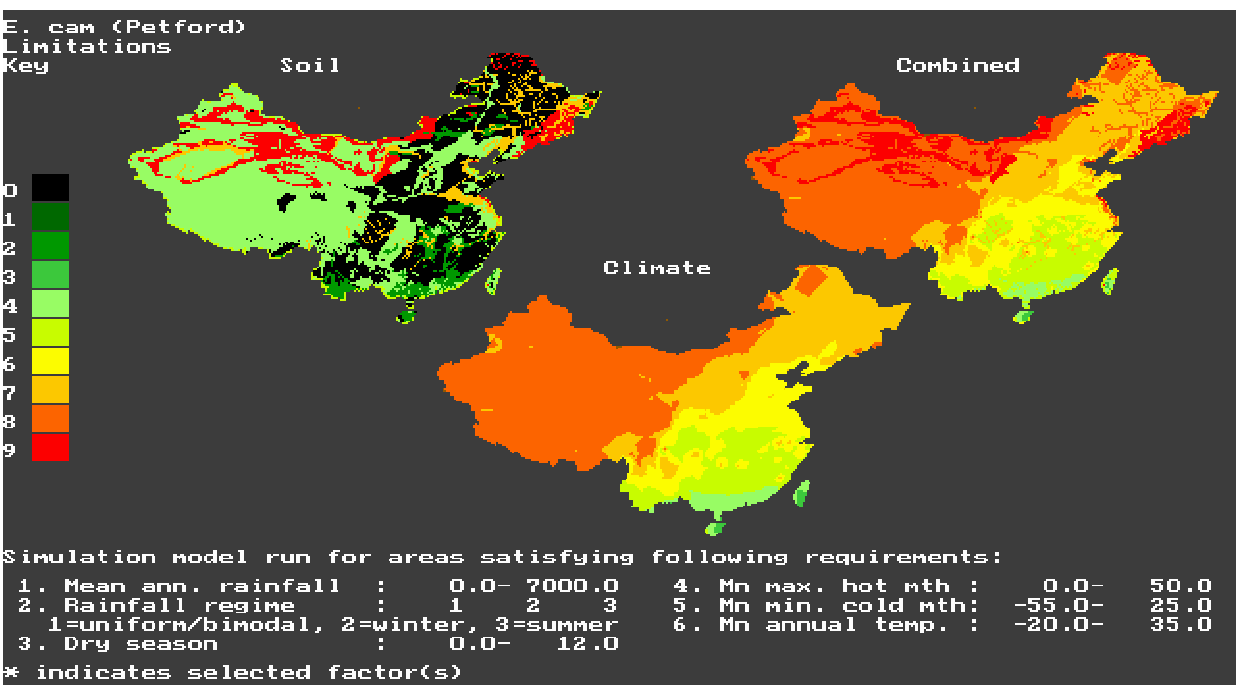
The Plantgro program has recently been used as part of a multi-million dollar project developing a 'National Masterplan for Forest Plantations' (NMFP) in Indonesia. This work was carried out by the DHV consultancy company and some of the work is outlined in the ACIAR proceedings. For example, Davidson (1996) describes how Plantgro plant files were developed for about 50 tree species. For the better-known species, such as *Tectona grandis* (teak) and *Acacia mangium* numerous trial results provided a good basis for the development of the notional relationships which described the trees' responses to environmental conditions. For the lesser-known trees the notional relationships were more educated guesses. Though the exact form of a response may not be known, there is usually some evidence for a species general preferences, for example, its need for acidic, neutral and/or alkaline soil conditions. Pawitan (1996) describes how a batch file version of Plantgro was developed, Plantgro limitation ratings were related to standard growth curves for different species to predict potential yield, yield predictions were used to evaluate the economic viability of particular projects and recommended land uses were plotted using the ARC/INFO and Arc/View GIS packages.

Plantgro was also used in ACIAR project 9127 and Hackett (1996) describes the use of foresters' expert opinions to develop Plantgro plant files. The development of a database of 244 Plantgro soil and climate files for Thailand is also described (Taweesak et al. 1996). The soil files were developed from detailed soil chemical and physical analyses, which had been carried out for four horizons at sites representing the major soil types of Thailand. The climate files were prepared using the interpolation relationships for Thailand (Zuo et al. 1996). Using the tree files developed for the NMFP project it would be possible to estimate the potential productivity of 50 tree species for all these 244 sites. As part of ACIAR project 9127 tree growth, soil and climate measurements were also collected from over 200 tree trial plots in Thailand and China (Sirirat et al. 1996, Yan Hong 1996).

Simulation mapping programs

The Plantgro program is generally used to predict growth at individual locations or small numbers of sites. However, it is useful to be able to see the suitability for particular species and provenances over wide areas both to check descriptions of requirements and to make recommendations of trees for particular regions. PC-based simulation mapping programs allow a simplified version of the Plantgro model to be run for thousands of grid points. They use interpolated monthly mean climatic data to carry out the same light, non-linear heat sum and water balance calculations as Plantgro, but use information from maps to carry out a simplified assessment of soil limitations.

The ACIAR proceedings includes a description of simulation mapping programs developed for Africa, Australia, Thailand and China containing data for 10 187, 11 299, 6 242 and 15 789 grid points respectively (Booth 1996c). The programs output three maps showing limitations for soils, climate and the two factors combined. For example, Figure 3 shows the suitability of 15 789 sites across China for the Petford provenance of *Eucalyptus camaldulensis*. In the example shown the model was run for all 15 789 grid points, a process which takes about 85 seconds on a 90mhz Pentium PC. It is possible to speed the operation of the program by restricting the analysis to areas satisfying a description of climatic requirements similar to those used by climatic mapping programs. This description is shown below the maps.



A marker can be moved over any location and a summary of the month-by-month limitations is shown for that particular location (see Figure 4).

Lat 23.25 Long 113.25 Elev 17

Soils (FAO type): 'Ge73-2/3a' 'n/a'

Climate (6 factors): 1799.0 3.0 2.0 31.0 9.0 21.5

	J	F	M	A	M	J	J	A	S	O	N	D
Soil	4	4	4	4	4	4	4	4	4	4	4	4
Solar	3	4	4	3	3	3	2	2	2	3	3	3
Temp	8	7	6	2	0	0	0	0	0	1	4	7
Water	4	3	1	0	0	0	0	0	0	0	0	2
Climate	8	7	6	3	3	3	2	2	2	3	4	7
Overall	8	7	6	4	4	4	4	4	4	4	4	7

Mean Limitations
Soil : 4
Climate : 4
Overall : 5

Press any key to return to maps

In Figure 4 the Ge73-2/3a soil type was assessed as having a level 4 limitation rating, which remains unchanged for all the months of the year. In January limitations for solar radiation, temperature and moisture were 3, 8 and 4 respectively. Applying Liebig's Law of the Minimum the greatest limitation in January is due to temperature and is rated as 8 (i.e. a major limitation). In contrast the greatest limitation in July is due to soil factors and is a moderate rating of 4. The colours shown for this single location on each of the three maps in Figure 3 simply indicate the mean limitations for soil, climate and overall conditions.

Discussion

The need for effective integration of GIS and environmental modelling is probably greater in developing countries than in the rest of the world, as environmental systems in many areas are already under great strain. At the same time the support available for sophisticated technologies is a fraction of that available in wealthy nations. Appropriate technologies need to be quick, simple and cheap.

Reliable climatic data are essential for predicting plant growth and modern interpolation methods can provide this information economically for any location on earth. Climatic mapping programs provide an effective means of delivering this information to users with minimal computing facilities and GIS skills. The Plantgro model provides a simple means of providing estimates of the potential growth of lesser-known plants. It was originally designed to work if necessary on text-only microcomputers, but the National Masterplan for Forest Plantations project has shown that it can also play a vital part in a multi-million dollar GIS analysis. Simulation mapping programs provide broadscale Plantgro-based analyses on PC-based systems.

Preparing appropriate tools and making them available either free or at minimal cost is only part of the process of ensuring the uptake of environmental assessment methods in developing countries. Training has been an important part of the work described here. In the last three years courses have been given in Thailand (2), China (2), Vietnam, Indonesia and the Philippines.

Generally, one week training courses have been provided. The first day has been an open seminar, which could be attended not only by the main group of trainees, but also by senior decision-makers and students (see, for example, Murdiyarso and Booth 1994). In the following three or four days a much smaller group of 12-16 trainees have been given "hands-on" training in the use of the climatic mapping, simulation mapping and Plantgro programs, usually operating two to a computer.

A course in Vietnam was successfully given in 1994 using microcomputers with 286-type processors. These modest machines were not only capable of running the programs described here, but were also used to show an animated fly-by of a digital elevation model of Vietnam (Booth 1995). Animations are useful in providing a quick appreciation of the effects of topography on climate. A good example of the effective uptake of the training provided in Vietnam was the use of a climatic mapping program by one of these trainees after the course to develop descriptions of the climatic requirements of nine native and nine exotic tree species important for plantations in Vietnam (Nghia 1996). Maps generated using climatic mapping programs are also beginning to appear in reference texts, such as "Growing Exotic Trees in China" (Pan and You 1994) and "Trees for Saltland" (Marcar et al. 1995).

Validation is a major problem with developing and applying models such as Plantgro. Some brief reports of validation work are included in the ACIAR proceedings (Booth 1996). However, more validation work needs to be done particularly with large datasets (i.e. >50 sites). Unfortunately, few datasets include information in sufficient detail or from a large enough number of sites to provide a really effective basis for validation (e.g. Schonau 1969, Hunter and Gibson 1984). Even where such large datasets exist, access may be restricted for commercial reasons. The large datasets which do exist tend to be from single countries and therefore usually do not explore the full range of conditions under which a species may be grown. Opportunities to develop large datasets by combining information from several countries are severely restricted because of the lack of an internationally agreed minimum dataset for recording results from forestry trials. Forestry trials are expense to establish and maintain, so it is unfortunate that greater efforts are not made to encourage the sharing of information. There is a great need for organisations such as the Center for International Forestry Research (CIFOR) to establish standards which would facilitate the exchange of data between countries.

The programs described here are of great help in assisting species introductions. However, the decision to introduce new species to an area should not be taken lightly and ecological as well as socio-economic impacts need to be carefully considered. Small scale trials should always be undertaken before large scale plantation establishment is attempted. Attention should also be given to establishing plantations which are ecologically sustainable (Nambiar and Brown 1996).

More information about the methods described here is included in ACIAR Proceedings no. 63 'Matching Trees and Sites' which is available from Bibliotech, GPO Box 4, Canberra, ACT 2601, Australia (fax +61 6 257 5088). Persons and institutions working in relevant areas in developing countries who may be eligible for a free copy should write to Publications, ACIAR, GPO Box 1571 Canberra, ACT 2601, Australia.

Acknowledgements

I am grateful to the Australian Centre for International Agricultural Research (ACIAR) and the Australian Agency for international Development (AusAID) for their financial support of the work described in this paper. I am also grateful to numerous collaborators in Australia, China, Thailand, Vietnam, Indonesia, the Philippines, Laos, Costa Rica, Colombia, Kenya and many other countries. Full acknowledgements are given in several papers included in the 'Matching Trees and Sites' publication (Booth 1996).

References

Booth, T.H. (1985) A new method for assisting species selection. *Commonwealth Forestry Review* 64: 241-250.

Booth, T.H. (1990) Mapping regions climatically suitable for particular tree species at the global scale. *Forest Ecology and Management* 36: 47-60.

Booth, T.H. (1995) Flying around the world. *GIS User* 14: 18-20.

Booth, T.H. ed. (1996a) Matching Trees and Sites. *Proceedings of an international workshop held at Bangkok, Thailand, 27-30 March 1995. ACIAR Proceedings No. 63.*

Booth, T.H. (1996b) The development of climatic mapping programs and climatic mapping in Australia. In Booth, T.H. ed. Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Booth, T.H. (1996c) Simulation mapping programs for Africa, China, Thailand and Australia. In Booth, T.H. ed. Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Booth, T.H. and Pryor, L.D. (1991) Climatic requirements of some commercially important eucalypt species. *Forest Ecology and Management* 43: 47-60.

Busby, J.R. (1991) BIOCLIM - a bioclimatic analysis and prediction system. In Margules, C.R. and Austin, M.P. eds. *Nature Conservation: cost effective biological surveys and data analysis.* Melbourne: CSIRO, pp. 64-68.

Davidson, J. (1996) Developing Plantgro plant files for forest trees. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Eastman, J.R. (1993) *IDRISI version 4.1.* Clark University, Graduate School of Geography, Worcester, Massachusetts, 209 pp.

ESRI (1992) *ArcView User's Guide.* Environmental Systems Research Institute, Redlands, 164 p.

Fryer, J. (1996) Site sampling and growth prediction in Central America. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Godwin, D.C., Ritchie, J.T., Singh, U. and Hunt, L. (1989) *A User's Guide to CERES-Wheat v 2.10.* Muscle Shoals, Alabama 35662, USA: International Fertilizer Development Center, 86 pp.

Hackett, C. (1988) *Matching Plants and Land: Development of a broadscale system from a crop project for Papua New Guinea.* CSIRO Division of Water and Land Resources. Natural Resources Series no. 11, Melbourne, 82 pp.

Hackett, C. (1991) *Plantgro : a software package for the coarse prediction of plant growth.*

Melbourne: CSIRO

Hackett, C. (1996) A study of forest scientists perceptions of trees' environmental relationships : implications for predicting growth. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Hutchinson, M.F. (1989). *A new objective method for spatial interpolation of meteorological variables from irregular networks applied to the estimation of monthly mean solar radiation, temperature, precipitation and windrun.* CSIRO Division of Water and Land Resources, Tech. Memo. 89/5, Canberra: CSIRO, 10 p.

Hutchinson, M.F. (1992) *Documentation for SPLINA and SPLINB - two programs in the ANUSPLIN software package.* Canberra: CRES, Australian National University.

Hunter, I.R. and Gibson, A. R. (1984) Predicting *Pinus radiata* site index from environmental variables. *N.Z. Journal of Forestry Science* 14: 53-64.

Jovanovic, T. and Booth, T.H. (1996a) The development of interpolated temperature and precipitation relationships for the Indonesian Archipelago. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Jovanovic, T. and Booth, T.H. (1996b) The development of climatic interpolation relationships for the Philippines. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Liebig, J. (1855) *Die Grundsatze der Agricularchemie mit Rucksicht die in England angestellten Untersuchungen.* Braunschweig: F. Vieweg und Sohn,

McMurtrie, R.E., Landsberg, J.J. and Linder, S. (1989) Research priorities in filed experiments on fast-growing tree plantations: implications of a mathematical production model. In Pereira, J.S. and Landsberg, J.J. (eds.) *Biomass Production by Fast-growing Trees.* London: Dortrecht, pp. 187-207.

Marcar, N., Crawford, D., Leppert, P., Jovanovic, T., Floyd, R. and Farrow, R. (1995) *Trees for Saltland : a guide to selecting native trees for Australia* East Melbourne: CSIRO.

Murdiyarso, D. and Booth, T.H. (1994) Evaluation of Climatic and Soil Data for Agriculture, Forestry and Conservation. *Proceedings of a Seminar and Workshop. Bogor Agricultural University* , 113 pp.

Nambiar, E.K.S. and Brown, A.G. (1996) *Management of Soils, Nutrients and Water in Tropical Plantations.* Canberra: Australian Centre for International Agricultural Research.

Nghia, Nguyen Hoang (1996) Climatic requirements of some of the main tree plantation species in Vietnam. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. In Longmore, R. ed. *Atlas of Australian Elapid Snakes.* Bureau of Flora and Fauna, Canberra, ACT, 4-15.

Pan Zhigang and You Yintian ed. (1994) *Growing Exotic Trees in China.* Beijing: Beijing Sci. and Tech. Press, 756 p.

Pawitan, H. (1996) The use of Plantgro in forest plantation planning in Indonesia. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Schonau, A.P.G. (1969) A site evaluation study in black wattle (*Acacia mearnsii* De Wild.) *Annale Universiteit van Stellenbosch* 44: 79-214.

Singh, U., Ritchie, J.T., and Godwin, D.C. (1993) *A User's Guide to CERES-Rice v 2.10* Muscle Shoals, Alabama 35662, USA: International Fertilizer Development Center, 130 pp.

Sirirat Janmahasatien, Chingchai Viriyabuncha and Snowdon, P. (1996) Soil sampling and growth prediction in Thailand. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Taweesak Vearasilp, Jovanovic, T. and Booth, T.H. (1996) Plantgro soil and climate database for Thailand. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Webb, D.B, Wood, P.J, and Smith, J.P. (1980) *A guide to species selection for tropical and sub-tropical plantations.* Oxford: Commonw. For. Inst. Oxford, Trop. For. Pap. 15, 342 pp.

Webb, D.B, Wood, P.J, Smith, J.P. and Henman, G.S. (1984) *A guide to species selection for tropical and sub-tropical plantations.* Oxford: Commonw. For. Inst. Oxford, Trop. For. Pap. 15, 256 pp.

Yan Hong (1996) Site/genotype matching and growth prediction for Australian trees in China. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Zuo, H., Hutchinson, M.F., McMahon, J.P. and Nix, H.A. 1996. Developing a mean monthly climatic database for China and Southeast Asia. In Booth, T.H. (ed.) Matching Trees and Sites, *ACIAR Proceedings No. 63.*

Trevor H. Booth
CSIRO Division of Forestry
PO Box 4008
Queen Victoria Terrace
Canberra
ACT 2600
Australia

Telephone : +616 281 8259
Fax : +616 281 8312
email : Trevor.Booth@cbr.for.csiro.au

From Third International Conference/Workshop on Integrating GIS and Environmental Modelling, Santa Fe, New Mexico, 21-25 January 1995, Proceedings published on WWW and CDROM by National Center for Geographic Information and Analysis, Santa Barbara.

DEFORESTATION IN TWO BRAZILIAN AMAZON COLONIES: ANALYSIS COMBINING FARMER INTERVIEWS AND GIS

Sam Fujisaka 1 Tel.(57)-2-4450-000; Fax.(57)-2-4450-273; e-mail. S.Fujisaka@cgnet.com,
Nick Thomas & Euan Crawford

Centro Internacional de Agricultura Tropical (CIAT), A.A. 6713, Cali, Colombia

---

ABSTRACT

Tropical deforestation is highest in Latin America compared to Africa and Asia. Settlers were interviewed in two Amazon colonies about land use and rates and causes of deforestation. Farmers in Pedro Peixoto, Acre, cleared about two ha per year per family and settlers in Theobroma, Rondonia, cleared some three ha per year to produce first rice followed by beans, maize, and cassava. Settlers then converted lands to pasture not only to raise cattle, but also as a way to add substantial value to lands for "improvements"--i.e., for more clearing, pasture, fencing, corrals and ponds. GIS analysis of satellite images and cadastral maps of the two colonies provided invaluable data on the dynamics of deforestation--i.e., on deforestation patterns and rates as influenced by distance to roads, wet-season access, land tenure, parcel size, and rates over time.

---

1. INTRODUCTION

Tropical deforestation, due in part to slash-and-burn agriculture, contributes to global warming via burning and release of CO_2 into the atmosphere; and Brazil is now the fourth atmospheric carbon contributor--after the US, ex-Soviet Union, and China (Moran 1993). Deforestation is also leading to losses of genetic (Phillips *et al* 1994) and cultural diversity. Decreasing transpiration and precipitation within and outside of areas cleared may also be a consequence of deforestation (Fearnside 1985, Salati 1989).

Farmer settlers in the government colonization projects of Pedro Peixoto in the Brazilian Amazonian state of Acre and in Theobroma, Rondonia (Figure 1), were interviewed and GIS analysis was conducted as a part of activities to characterize local land use systems and the dynamics of deforestation. Settlement in such projects has been facilitated by government policies to populate frontier areas in the Amazon, road construction, and direct and indirect subsidies. This paper addresses dynamics of deforestation at the farm level.

Figure 1. Location of Pedro Peixoto, Acre, & Theobroma, Rondonia, in the Brazilian Amazon

The Pedro Peixoto colonisation site, established in 1972, covers 370,000 ha divided into 3700 lots distributed among 3200 families. Lots are located between 50 and 100 km from the state capital of Rio Branco, with the major highway BR364 passing from Rio Branco through the settlement and on to Porto Vehlo in Rondonia.

Theobroma, established in 1979 (albeit spontaneously settled much earlier), officially covers 300,000 ha divided into 3000 lots (reportedly) distributed among 3000 families. The project area is located some 350 km from the state capital of Porto Vehlo, also via BR364 which connects Porto Vehlo and Cuiaba in the south (D'Oliveira, unpublished)

2. METHODS

Farmer-settler interviews. A draft questionnaire was prepared by researchers from CIAT, the Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA), the International Centre for Research in Agroforestry (ICRAF), and the International Food Policy Research Institute as a first step in a series of characterization research activities, all, in turn, a part of the ICRAF-coordinated project "Alternatives to Slash-and-Burn" (ASB). The questionnaire was field tested and further modified by CIAT, IFPRI, and EMBRAPA researchers to ensure appropriateness to local conditions (similar work being conducted in Cameroon and Indonesia) and to facilitate ease of use in the field and data coding. Eighty-one randomly sampled farmers in Pedro Peixoto and 74 in Theobroma were interviewed in late August and early September 1994. Either or both male and female heads of household were included. We reviewed and cross-checked data, revised the coding system as needed, set up data archives, and entered and tabulated data.

Image selection. Landsat Thematic Mapper (TM) imagery was used in this study. Selection of imagery considered project goals, availabilility, and cost. Unlike SPOT (Multi Spectral, Panchromatic) data, Thematic Mapper, Multi Spectral Scanner (MSS), and ERS1 radar could accomodate either project area within a single scene. Aerial photographs were eliminated from consideration due to the substantial labor required to georeference and mosaic them. Use of MSS data--although available as part of a large historical archive--was discounted due to its spatial resolution of approximately 70 m2 because farmer's parcels ranged between 200 and 450 m in width; and data would not have afforded the detail of landcover desired. Radar imagery was not considered due to the lack of an historical archive.

Scenes were acquired from two sources. Three Pedro Peixoto images from 1984, 1987 and 1992 were obtained from the Pathfinder project at the Universities of Maryland and New Hampshire. The single TM image (1994) of Theobroma was purchased from the Brazilian Instituto Nacional De Pesquisas Espaciais (INPE). As with the Pedro Peixoto images, this was recorded in the dry season.

Cadastral maps. Digital cadastral maps of both areas were obtained from the Instituto Nacional de Colonizacao e Reforma Agraria (INCRA). Lot boundaries, partial road networks, population centres and drainage systems were present in both cadastral maps. The Pedro Peixoto coverage also indicated lots whose owners had obtained land titles prior to 1992. The cadastral maps were georeferenced to a UTM projection.

Ground truthing. Fieldwork was conducted in Pedro Peixoto and Theobroma. We interviewed individual farmers regarding their land management practices. Landcover samples were taken by interviewing farmers and walking lots with a handheld GPS (+/-100m) while making field sketches.

Image preprocessing. An atmospheric correction (dark pixel subtraction method) was applied to all images to provide comparable classifications. Occasional line dropout was removed using a filter targeted at spurious lines. All images were georeferenced to the cadastral overlay for two reasons. First, topographic map coverage at the 1:100,000 scale was available but had not been updated since 1977. Second, as we wanted analysis at the parcel level it was more important that the imagery was accurately referenced to the cadastral lots than to the national projection.

The number of ground control points (GCPs) used in the correction varied according to relief. Pedro Peixoto, an area of low relief, was georeferenced using 75 GCPs, giving a Root Mean Square (RMS) error of less than one. Due to its slightly more undulating terrain, Theobroma was corrected using 90 GCPs, again to an RMS error of less than one.

Classifications. To quantify rates of deforestation the project required a forest/non-forest classification. Non-forest included grazed and ungrazed pasture, agricultural plots and secondary regrowth. Forest included only forest (a category which included "disturbed" forest). All images were classified to the 95% threshold of assignment using isodata clustering. Twenty five cover classes were created and visually interpreted using the ground truth data into either forest/non-forest. This allowed the study of deforestation over time in Pedro Peixoto. As the Theobroma study required measures of secondary forest regrowth, the non-forest category was split into secondary forest regrowth and land currently worked by the colonist.

A majority filter was run on all classified images to improve accuracy.

3. RESULTS

Land use from farmer interviews. A main research objective was to determine land use and patterns of land use change. The national government granted settlers large parcels of forested land which they cleared for agricultural use, starting with slash-and-burn agriculture. Parcels were a mean 88 ha in Pedro Peixoto and 76 ha in Theobroma. By 1993-94, settlers had cleared a mean 27 ha (31%) of these lands in Pedro Peixoto and 35 ha (46%) in Theobroma (difference not significant at 5% using the t-test). The 31% of cleared land in Pedro Peixoto were divided into a mean 20% (of the total parcel) in pasture, 6% in fallow, and 4% in annual crops. The 46% of cleared lands in Theobroma were divided into 26% pasture, 8% fallow, 7% annual crops, and 5% perennial crops (Table 1).

Changes in land use from 1993-94 to 1994-95 could be calculated because interviews were conducted in late 1994 after field clearing and burning as farmers prepared for the 1994-95 cropping season: deforested portions of the settlers' parcels increased a mean 2.0 ha and from 31% to 34% in Pedro Peixoto and a mean 2.7 ha and from 46% to 50% in Theobroma. Overall, some 40% of the settlers' land in the two colonies has been deforested, with more than half of the cleared area converted to pasture. Only 7% of settlers' lands were in fallow (Table 1).

Table 1.

Land use (mean areas), Pedro Peixoto, Acre (n=81) & Theobroma, Rondonia (n=74), 1993/94 & 1994/95

Tabla 1.

	Pedro Peixoto				Theobroma				
	93/94		94/95	Dif	93/94		94/95	Dif	
	ha	**%**	**%**	**%**	**ha**	**%**	**%**	**%**	**Total**
Forest	61	**69**	66	-5	41	**54**	50	-7	**61**
Cleared	27	**31**	34	+11	35	**46**	50	+8	**39**
Pasture	**17**	**20**	**25**	**+30**	**20**	**26**	**29**	**+10**	**23**
Fallow	5	6	2	-60	6	8	4	-50	7
Annual crops	4	4	7	+50	5	7	9	+40	6
TOTAL	**88**	100	100		**76**	100	100		100

The interviewed farmers had converted a mean 19 ha in Pedro Peixoto and 30 ha in Theobroma of primary forest at the time of the interviews. A mean 3.3 ha in Pedro Peixoto and 8.5 ha in Theobroma had been cleared at the time of arrival and parcel occupation. As indicated above, farmers had occupied parcels in Pedro Peixoto a mean 9 years and Theobroma a mean 8 years. The rate of primary forest clearing was thus calculated as 1.8 ha per year in Pedro Peixoto and 2.8 ha per year in Theobroma--figures very close to those calculated for 1994-95.

In both colonies, individual farmers' areas cleared and area in pasture were correlated to overall parcel size; and area in pasture was highly correlated to area cleared. For Theobroma, farmers who did not clear forest land in late 1994 had significantly (at the 5% level) more fallow land (8.6 ha) than those clearing (4.0 ha), suggesting fallow use by some Theobroma farmers. Although Pedro Peixoto farmers clearing forest in late 1994 also had less fallow land (5.6 ha) than those not clearing (4.5 ha), the difference was not significant.

Rice was the major crop for both consumption and sales of surpluses at both sites. Farmers (92% in Pedro Peixoto and 70% in Theobroma) planted rice in the first year of cultivation of what was primary forest; and cultivated maize, cassava (in Pedro Peixoto), and pasture in the second year. Rice was not grown in the second or subsequent years of plot use.

Farmers cultivated lands cleared from primary forest for a mean 2.1 (Pedro Peixoto) to 2.5 years (Theobroma). Sixty percent in Pedro Peixoto cultivated such plots for two years; while Theobroma farmers used their newly cleared lands for from one to more than three years in somewhat equal proportions. Farmers reported that discontinuation of annual cropping on lands cleared from forest was due to the not mutually exclusive reasons of lower productivity, weeds--especially *Imperata* sp (locally *sape*), and insects and diseases.

After food crops, two-thirds of Pedro Peixoto and nearly half of Theobroma farmers converted their lands to pasture. About a third in both areas left some land in fallow (although much of the "fallow" could also serve as unimproved pasture). Theobroma (20%) but not Pedro Peixoto farmers also converted some land from annual to perennial crop use. Farmers at both sites "normally" left any fields which they fallowed for a mean 2.5 years, although they thought that 3.0-3.5 years of fallow would be ideal. Rice followed by maize and beans were the main crops planted in re-opened fallows.

Cattle and pasture formation are perhaps the major driving force behind deforestation in the settlements. Most settlers (91% in Pedro Peixoto and 81% in Theobroma) had cattle. Herd size was a mean 18 head (with 6 giving milk) in Pedro Peixoto and 26 (4 giving milk) in Theobroma.

Settlers' main source of wealth appears to be the appreciation of the value of their lands due to conversion to pasture: 93% of Pedro Peixoto and 97% of Theobroma settlers perceived their land values as having risen (values were discussed in terms of equivalent numbers of cattle), at annual rates of 74% in the former and 157% in the latter site. Farmers reported total increases since occupying their parcels in value of about 800% in Pedro Peixoto and 950% in Theobroma, with main reasons for increases attributed to addition of pasture or cleared areas, fencing, ponds, and corrals (Table 2).

Table 2.

Respondents' evaluation of and reasons for (% of respondents) changing land values, Pedro Peixoto (n=69) and Theobroma (n=70)

Table 2.		
	Pedro Peixoto	**Theobroma**
Report increased value (%)*	93	97
Total mean increment in value (%)	778	952
Mean annual increment in cattle (%)	74	157
Reasons for increase		
More pasture	60	50
Fencing	56	36
Pond/water	30	13
More cleared area	12	26
Corral	12	16
House	26	16
Access/roads	25	27
Perennial crops	12	35
School	10	5
Timber	8	1
Title	7	3
Good soils	4	5

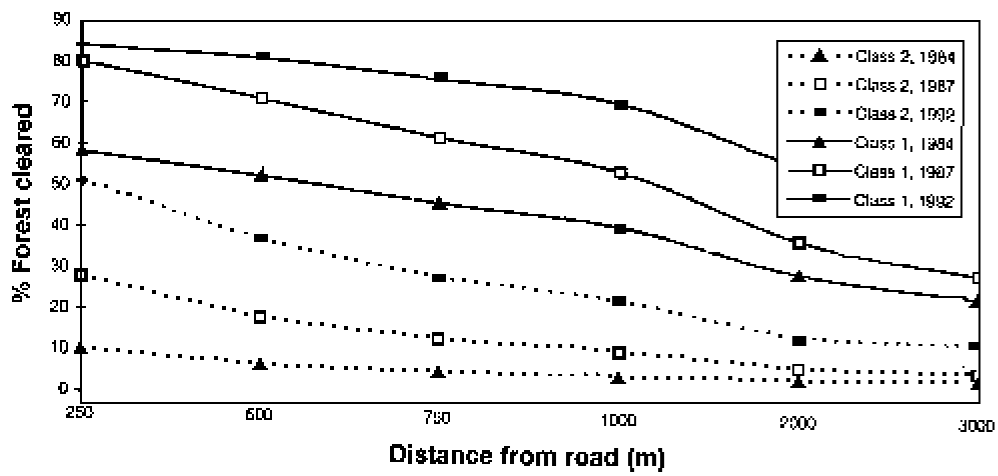
***** calculated in terms of numbers of cattle at time of parcel acquisition and at time of interview

---

GIS: distance to roads and deforestation, Pedro Peixoto. Accessibility of areas has been shown to play an important role in the spatial distribution of deforestation. "Accessibility" in this case was defined in terms of distances to roads and farmers' evaluations of wet-season inaccessibility. No modern road network map existed for Pedro Peixoto. A partial road network was extracted from the cadastral map. The "missing links" were digitised to an accuracy of +/- 30 metres using an edge-enhanced georeferenced TM image, expert knowledge, and the cadastral map. Roads were classified into two categories, the major highway--the trans Amazonian BR364--and project roads, which eventually connect to highway. Distance was measured by buffering the roads. A preliminary study was conducted to assess the optimal buffer distance, a compromise between a high level of detail and a manageable amount of data to process. Two hundred and fifty metres was selected as optimal. Buffering was conducted across the whole study area with the exception of road junctions. Confusion created by proximity to both road classes exists at these locations. Deforestation rates were calculated within the GIS which held both the vector buffer zones and the classified raster images.

From 1984 to 1992, deforestaion increased: a) from 58% to 84% for areas up to 250m away from the main roads; b) from 25% to 48% at distances between two and three kilometres from the main road; an c) at intermediate levels for the intermediate distances. For secondary roads, deforestation similarly increased over the 1984-92 period from 10% to 51% for the area up to 250 m from the road; and from 4% to 13% over the same period and at two to three kilometres from the secondary roads. Intermediate distances (and the intermediate 1987 image) from the secondary roads again provided intermediate values (Figure 2).

Figure 2. Forest Clearing as a function of distance from roads, Pedro Peixoto, 1984, 1987 and 1992



Class 1 are paved highways, class 2 are access roads within the colony

GIS: wet-season access and deforestation, Pedro Peixoto. Two hundred and fifty lots were randomly selected for accessibility analysis. With the aid of the Empresa Brasilera de Pesquisa Agropecuario (EMBRAPA) the lots were categorised into those accessible and those inaccessible in the wet season. Again, rates of deforestation by lot type were calculated using the GIS. Deforestation was initially lower in less accessible compared to more accessible lots.

In 1984, 5% of the inaccessible compared to 9% of the accessible lots were deforested. By 1992, however, the gap had narrowed to an insignificant difference of 25% of inaccessible vs 27% of accessible lots (Table 3).

Table 3.

Deforestation & wet-season access, Pedro Peixoto

Table 3		
	1984	**1992**
Accessible (mean & deforestation)	9	27
Not accessible	5	25

GIS: land tenure and deforestation, Pedro Peixoto. Data regarding parcel ownership was available for 1992. We hypothesized that land ownership has an effect--positive or negative--on deforestation rates. Three hundred and twenty four lots were randomly selected, 160 with titles and 164 without. The 1992 forest/non-forest classification was vectorised and unioned with the cadastral map to give the forest/non-forest boundaries within sampled lots. This coverage was interrogated to give forest/non-forest areas in hectares for each sample lot. Data was assimilated for lots with and without title to give actual hectarages and landcover as a percentage of total lot area.

Deforestation amounts were converted to percentages as a function of lot area. Lots were sorted into percentiles of deforestion according to ownership (Table 4). Differences in levels of deforestation were significant at the 99.9% confidence level for classes of less than 30%, between 30 and 49.9% and over 50% deforestation. Colonists with title by 1992 were clearing a greater percentage of forest than those without. This finding is somewhat contrary to the idea that secure land tenure results in more careful resource management.

Table 4.

Deforestation x land tenure, Pedro Peixoto

Table 4			
	Deforestation		
Tenure	**> 30%**	**31 - 49 %**	**> 50%**
Have title (% of farmers)	46	34	20
Do not have title	65	27	9

GIS: parcel sizes and deforestation, Theobroma. The colonisation scheme is compromised of 100 ha and 50 ha lots. A hypothesis was that the area and extent of primary forest clearing per lot, and its replacement with both secondary regrowth and managed land, would not be affected by lot size. As each lot was occupied by a single family, it was assumed that the same workforce was available for agriculture and forest clearing. Two hundred and fifty nine lots were randomly selected: one hundred and twenty 100 ha lots and one hundred and thirty-nine 50 ha lots. The three-class--primary forest, secondary regrowth and managed lands--classification was converted into a vector coverage and unioned with the cadastral overlay of land parcel boundaries. This coverage was queried and values of areal extent and extent as a percentage of the lot area obtained.

Large lot holders had cleared significantly greater areas than small lot holders; absolute amounts of secondary regrowth were greater for large lots; and amounts of cultivated and pasture land in the large lots was significantly greater than that in small lots. Percents of lot areas under secondary regrowth for large and small lots, however, were not significantly different. Although large lot holders may have cleared their lands at a faster rate, we conclude that they have been clearing at the same rate as small holders, but started sooner. The situation is confounded by access: large lots are also located closer--i.e., are more accessible--to the main highway (Table 5).

Table 5.

Deforestation x parcel size, Theobroma

Table 5		
Land use	**Small (%)**	**Large (%)**
Primary forest	25.2(49)	32.4(31)
Secondary regrowth	7.6(16)*	18.4(18)*
Pasture/cultivated	17.2(36)	51.2(51)

***** All differences except indicated are significant at 0.01 using the student's-t

---

GIS: deforestation over time, Pedro Peixoto. The time series of classified Pedro Peixoto images was used to create a visual representation of deforestation. By using image subtraction a map indicating land cleared up to 1984, land cleared between 1984 and 1987, and land cleared between 1987 and 1992 was produced (Figure 3).

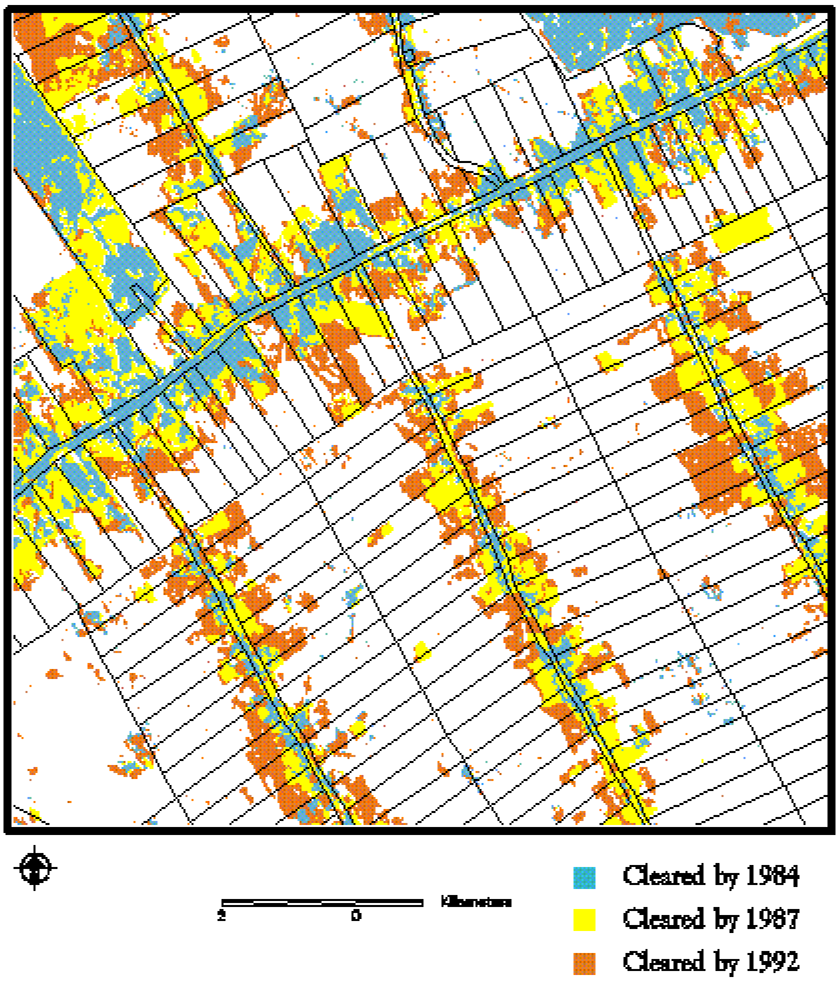
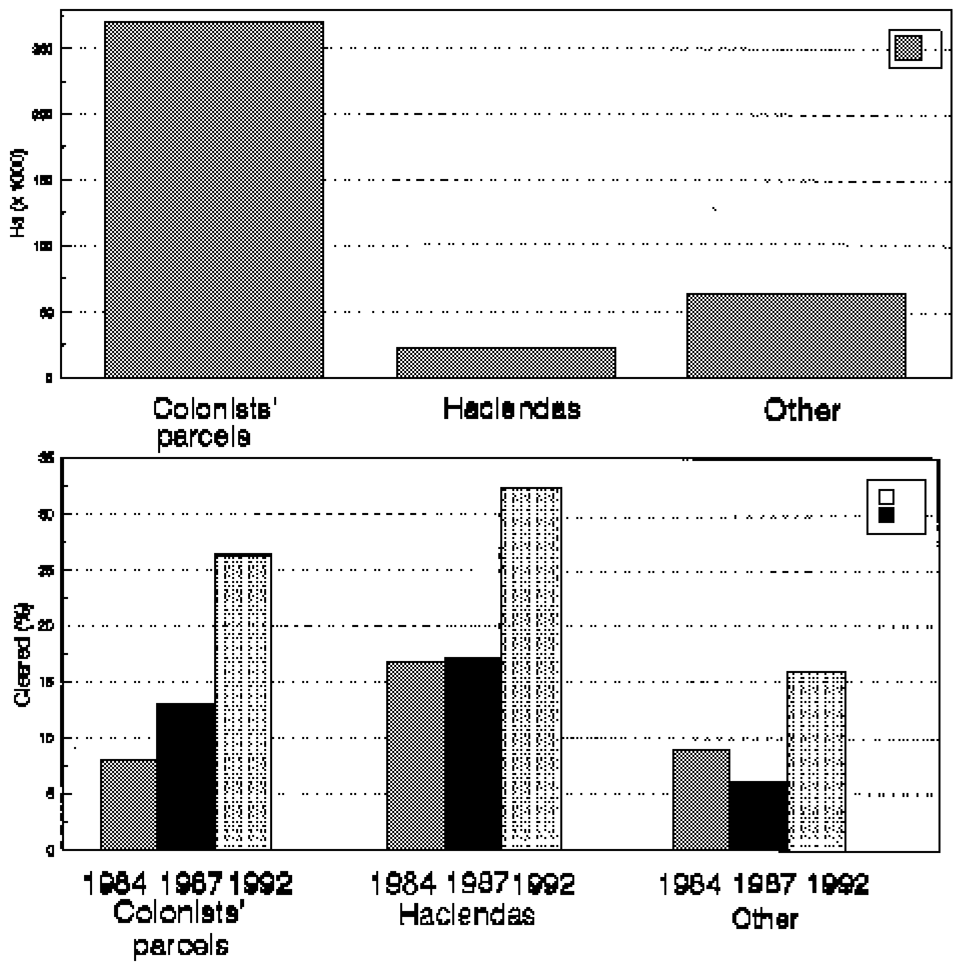


Figure 3 -Forest Clearing by lot, Pedro Peixoto, 1984, 1987 and 1992.

Analysis tends to confirm rates calculated from farmer-reported data. The images cover an area of 357,000 ha, of which there are 276,000 ha of colonists' parcels, 22,000 ha of haciendas, and 56,000 ha of "other" (the officially reported 370,000 ha of the Pedro Peixoto project thus appears to include haciendas and land uses besides settlers' parcels).

Overall, the percent area cleared was 8.8% in 1984, 12.1% in 1987, and 25.0% in 1992. Colonists' cleared areas increased from 8.1% in 1984, to 13.1% in 1987 and 26.4% cleared in 1992 (Table 6, Figure 4). This latter figure and the cleared area calculated from farmer interviews of 31% in 1993-94 are mutually supporting: both analysis provide annual rates of about 3% and total deforestation in Pedro Peixoto colonists' lots at about 30% in 1993.

Figure 4. Areas (ha) and Forest clearing (%) by major land uses (satellite image analysis) Pedro Peixoto, 1984,1987 and 1992



The large cattle ranches (*fazendas*) included in the images increased cleared area from 16.8% to 32.4% from 1984 to 1992. The area cleared on the ranches approximiately doubled over the eight year period (1984-92); while clearing on the colonists' parcels increased by more than 300% for the same period (Table 6).

Table 6.

Forest clearing (satellite image analysis), Pedro Peixoto, 1994, 1987, and 1992.

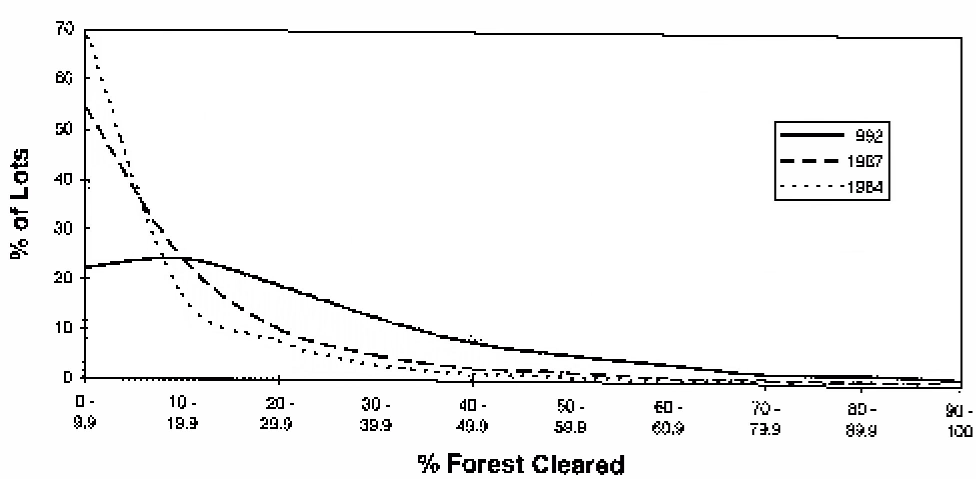
Table 6				
		% area cleared		
	Total ha	1984	1987	1992

Total (ha x 000)	356.7	8.8	12.1	25.0
Colonists' parcels	276.4	8.1	13.1	26.4
Haciendas	21.8	16.8	17.2	32.4
Other *	58.5	9.1	6.2	16.1
Accessible parcels **	79.5	9.3	13.7	27.1
Inaccessible parcels	169.9	5.3	11.7	24.6

For per parcel deforestation in Pedro Peixoto, images show that there was an increase from 22,388 ha cleared in 1984 to 72,970 ha cleared in 1992 for 3,141 farmer's parcels. That is, a total of 6,323 ha were cleared per year over the eight year period, or a rate of 2.0 ha cleared per parcel per year. This rate again corresponds to the calculation based on farmer-reported data.

Satellite image analysis was also used to determine the range and distribution of deforestation by deciles (i.e., frequency of lots showing 0-9.9%, 10-19.9%, 70% of the settlers' parcels were less than 10% deforested; and only 3% of the parcels were 40% or more cleared. By 1992 only 22% of the parcels remained less than 10% deforested; while another 22% were 40% or more deforested (Figure 5). The obvious and expected trend is that settlers' lots will steadily "move" towards the higher deforestation deciles and away from the lower. Modelling a date when, for example, no lots would remain in the less than 10% deforested and 10% of the parels would be deforested at the 90-100% level is complicated, however, by differential rates of forest clearing as a function of distances from roads and of access and other factors.

Figure 5. Deforestation by percentiles, Pedro Peixoto, 1984, 1987 and 1992



4. DISCUSSION

Data acquisition. The greatest difficulty was acquisition of satellite images. Very few cloudfree, or near cloudfree, images were available. Technical problems experienced by the South American receiving stations also dramatically reduced availability of recent imagery. As a result, fieldwork was conducted before the most recent images were acquired; which in turn meant that optimal strategies could not be developed for landcover sampling and farmer interviews (e.g., in areas of high and low deforestation). The problems encountered in image acquisition suggest that for future work, where a long historical archive is not required, radar or the more flexible airborne scanner (ATM) may be appropriate.

5. CONCLUSIONS

Deforestation at the farm level averaged two ha per year per parcel in Pedro Peixoto and three ha per year in Theobroma. Reasons for the significantly higher rate of deforestation in Theobroma were not determined, although the contrasting possibilities that clearing accelerates as the settlements age (Theobroma having been settled earlier) or that colonists decrease rates over time (the interviewed settlers in Theobroma had lived in that settlement for less time than had the Pedro Peixoto settlers) will be investigated in the future.

Settlers still had more than half of their lands in forest and will most likely continue to slash, burn, cultivate, and convert more primary forest as their currently most (economically) viable option. Two main factors driving land clearing at the farm level were the need to produce food and incentives (i.e., in the sense of increased land values) to convert land into pasture. In terms of food production, farmers consumed and sold rice, and to a lesser extent, beans, maize, and cassava (or cassava meal). Rice cultivation may "drive" some deforestation in that although farmers planted cleared fields for up to three years, they could not--for technical reasons--sow rice other than in the first year after clearing.

Farmers were clearly motivated to convert lands cleared from forest into pasture because of real or at least perceived resulting increases in land values. Farmers not only maintained cattle as standing "bank accounts" and obtained cash from sales of animals and milk, but built savings by investing time and resources in pasture, fencing, corrals, and ponds. As observed at the two sites, local ranchers and urban-based speculators have purchased continuous blocks of colonists' parcels to form new ranches or to expand the size of adjacent ranches; and payments were much higher for cleared vs forested portions of parcels.

Finally, we conclude that the combination of GIS analysis with intensive fieldwork (i.e., farmer interviews) provided the best optimal mix of methods to understand the dynamics of deforesation and land use in the Brasilian Amazon colonies studied.

REFERENCES

Fearnside, P.M. 1985. Environmental change and deforestation in the Amazon. *In*: J. Hemming (Editor), *Man's Impacts on Forests and Rivers*. Manchester: Manchester University Press.

Moran, E.F., 1993. Deforestation and land use in the Brazilian Amazon. *Human Ecology* 21:1:1-21.

Phillips, O.L., Hall, P., Gentry, A.H., Sawyer, S.A. and Vasquez, R., 1994. Dynamics and species richness of tropical rain forests. *Proceedings of the National Academy of Sciences, USA*. 91:2805-2809.

Salati, E., 1989. Deforestation and climatic changes in the Amazon Basin. *In: Climate and Food Security*. Los Banos, Philippines: International Rice Research Institute and the American Association for the Advancement of Science.

Timothy W. Foresman
Helen V. Wiggins
Dana L. Porter
Department of Geography
University of Maryland Baltimore County
5401 Wilkens Avenue
Baltimore, Maryland 21228

Penny Masuoka
University of Maryland Baltimore County
NASA Goddard Space Flight Center Code 920.2
Greenbelt, Maryland 20771

William Acevedo
U.S. Geological Survey
NASA-Ames Research Center MS 242-4
Moffett Field, California 94035

Design and Documentation of a Baltimore-Washington Regional Spatial Database Testbed for Environmental Model Calibration and Verification

---

ABSTRACT

Recent efforts by scientists and managers to inventory, map, and model impacts of human activities on the environment have focused on land transformation and urbanization processes. To test the efficacy of any single model, algorithm or procedure which defines land transformation processes a standard database calibration reference resource is required. Therefore, a set of georeferenced, spatially structured and well documented data sets has been designed for the Baltimore-Washington Region as a test and evaluation resource for the community of environmental modelers and global change scientists.

Land transformation processes are being examined from a variety of perspectives and scales using a variety of indicator parameters and mensuration variables. Tools and techniques applied to land transformation assessments range from creation of simple population expansion maps to change detection calculations using remotely sensed satellite data. A variety of point and cell growth models have been applied to simulate the land transformation phenomenon. These activities have demonstrated the reality that urbanization and land transformation processes involve complex interacting variables.

A team of scientists are expanding the efforts of the USGS Human Impacts on Land Transformation (HILT) project to build an Internet accessible "collaboratory" containing quality controlled spatially referenced calibration and validation databases. The Baltimore-Washington Regional Testbed provides for the calibration, verification, and validation for multiple scalar, temporal, thematic, and spectral assessments or models. This design and documentation procedures for creating the Baltimore-Washington Regional "Collaboratory" are presented in relation to its use for environmental modeling applications.

INTRODUCTION

Urbanization can be described as a massive unplanned global experiment affecting increasingly large acreages of the Earth's surface (Alig and Healy 1987). Peter Vitousek (1994) described the ongoing land use/land cover change, along with increasing concentrations of carbon dioxide in the atmosphere and alterations in the global nitrogen cycle, as well-documented factors of concern for global change community. This massive change in land surface characteristics is just beginning to be studied by Earth system scientists in terms of ecological processes, atmospheric implications and micro and macro-climatic impacts. Questions regarding ecosystem structure and function along the urban-rural gradient have been raised are appropriate for the global change agenda (McDonnell and Pickett 1990). Recent investigations along the urban-rural gradient have provided new insights into the apparent impacts urbanization has introduced to stable ecological systems (Pouyat and McDonnell 1991; et al., 1994; et al., 1995). Conceptually, Earth scientists may agree that urbanization is a key determinant in the litany of ecosystem transition processes of interest to the environmental modeling community. However, a significant gap exists in the capabilities to address design and implementation of integrated modeling structures along the urban-rural gradient. Challenges of integrating and developing models to understand and predict processes of urbanization affecting environmental conditions requires an interdisciplined approach towards collaboration of the talents, modeling resources, and spatial data at local, regional, and global scales (Asrar and Dozier 1994, Blood 1994, Pickett and Cadenasso 1995). A collaborative testbed for the Baltimore-Washington region has been designed to meet many of these integrated environmental modeling challenges. The testbed (or Collaboratory) is a comprehensive set of spatial databases put together in a manner which reconstructs and represents the real world.

In partnership with the US. Geological Survey (USGS), the University of Maryland Baltimore County is compiling historic maps, demographic data, environmental parameters, and satellite images to map human-induced land transformations for the Baltimore-Washington region. This effort includes collaboration with the Bureau of the Census, the University of California at Santa Barbara, the Goddard Space Flight Center, and numerous other federal, state, local, and private institutions. This work builds upon earlier research by the USGS that documented urban development phenomena for the San Francisco Bay region. That effort, working under the premise that historic overviews of urban sprawl can be provide insight into future scenarios, used a geographic information system to compile and visualize historic perspectives from 1850 to 1990 (Kirkland et al. 1994). A methodology was developed to combine the information from a variety of sources into an integrated, multi-scale, and multi-resolution dataset. This methodology was expanded for the Baltimore-Washington effort to promote a variety of environmental, social, and economic models related to urban-rural dynamics. Contemporary analysis focuses on the use of remotely sensed data, existing digital land use data, digital census information, a variety of Earth science infrastructure data, such as Digital Line Graphs, Digital Elevation Models, and other key ancillary demographic information. The resulting database of temporal urban demographic changes, which forms the framework of the Baltimore-Washington testbed, provides an ideal source of information to calibrate and verify models for urban geographers, environmental scientists, and global change scientists.

Design and construction of large spatial databases for environmental modeling has remained topical for both the GIS and environmental modeling community as evidenced in part by the interest and content of the three international conferences sponsored by the National Center for Geographic Information and Analysis. Significant progress is being made by the modeling community for employing effective GIS entity-relationship-attribute schemas thus offering promise of improved GIS integration with environmental models in general. Object and feature-based schema are described by many modelers as a successful path for improving performance of environmental models relative to articulation with spatial databases (Guptill and Fegeas 1988, Raper and Livingstone 1995). Use of the temporal domain for modeling, requisite for change or trend analysis, poses additional challenges for integrating GIS and environmental models. Peuquet (1994) offers an overview of temporal data structure theory that indicates various avenues of approach are available for time series analysis. Raper and Livingstone (1995) offer object-oriented structures as an approach to including temporal dynamics in environmental spatial models. While alternative designs are beginning to incorporate temporal datasets as input to environmental models, contemporary modeling with GIS remains primarily focused on the application of time slices defining geographic entities either in raster or vector data structures (Kemp 1993, Kemp and Kowalczyk 1994, Mitasova et al. 1995, Farmer and Rycoft 1991). The Baltimore-Washington regional testbed is designed along the more general time slice database structures with emphasis on improved metadata documentation. While more conservative in terms of database development, this approach offers better calibration opportunities to modelers with the standard arc-node entity definition and relational attribute definitions. This paper details the decisions and steps that led to the creation of the Baltimore-Washington regional testbed and the environmental science application of the spatial database resources.

BACKGROUND

As part of the U.S. Global Change Research Program, the USGS initiated urban mapping research activities to understand the urban transition from a historical and multi-scale perspective appropriate for modeling and predicting regional patterns of urbanization into the future (Kirkland et al., 1994). The USGS Human-Induced Land Transformations (HILT) project initially involved mapping the growth of urban development for the San Francisco Bay area using archival topographic maps and Landsat satellite images to delineate changes in the urban extent over time (Bell et al. 1995). Visualization of the urban data maps using time-series animation resulted in an effective videotape presentation to both scientists and the public alike (Acevedo and Bell 1994). The modeling component of this effort involved developing a cellular automation urban growth model using the San Francisco Bay database (Clarke et al. 1996). Clarke (et al. 1995) developed the model by adapting a wildfire behavior, environmental model and derived calibration techniques applicable to the land cover and other urbanization growth barriers using the temporal urban database. Continued refinement of the model which includes examining multi-scale extensions and interrelationships of various urban parameters is reported in a separate paper by Clarke (1996) at this conference.

Building upon the San Francisco Bay HILT activities was a collaboration that included UMBC, the U.S. Census Bureau and others to extend the temporal mapping into the Chesapeake Bay region. This collaboration created a multi-phased research plan, Table 1, which entails the creation of a multi-thematic, multi-temporal, multi-scale and multi-resolution spatial database structure for the greater Baltimore-Washington region, Figure 1.

Table 1. Baltimore-Washington Regional Collaboratory Projects

Themes and Activities	Phase I Baltimore 15-Minute Quadrangle Completed	Phase II Baltimore/Washington 2 degree x 2 degree Area Planned	Phase III Baltimore 15-Minute Quadrangle Planned	Phase IV Baltimore/Washington 2 degree x 2 degree Area Planned
Built Up Area-Temporal	MAC	MAC	Hunter	Hunter
Transportation-Temporal	MAC	MAC		
Coastline Temporal	UMBC/ MAC	UMBC	UMBC/ MAC	UMBC/ MAC
Population Density-Temporal	UMBC/Census	UMBC/ Census Hunter/UCSB	UMBC/Hunter UCSB	UMBC/Hunter UCSB
Bathymetry	DNR	DNR		
Wetlands-Temporal	DNR	DNR/GSFC	UMBC/DNR, UCSB/GSFC, MAC, and EDC	UMBC/DNR, UCSB/GSFC, MAC, and EDC
Land Use Land Cover MAGI	UMBC/DNR	UMBC	UMBC/DNR	UMBC/DNR
Economics-Temporal -County Districts-	UMBC	UMBC	UMBC/ Census	UMBC/Census
Political Boundaries Temporal	MAC/UMBC	MAC/UMBC		
Topography	UMBC	UMBC		
Science Modeling Applications	UMBC GSFC	UMBC/Yale/Hunter GSFC/UCSB	UMBC/Yale/Hunter GSFC/UCSB	UMBC/Yale/Hunter GSFC/UCSB
Statistical Applications	NMD/RES	NMD/RES	NMD/RES	NMD/RES
Visualization	EDC	UMBC NASA/GSFC	UMBC NASA/GSFC	UMBC NASA/GSFC

The Participants: Cooperative members listed above define lead role functions only for projects.

MAC- U.S.G.S. Mapping Application Center
EDC- Eros Data Center at Moffett Field
RES- U.S.G.S. Research Branch
GSFC/NASA- Goddard Space Flight Center
Yale - Yale School of Forestry and Institute of Ecosystems Study
UMBC- University of Maryland Baltimore County
Census- U.S. Bureau of the Census
NMD- U.S.G.S National Mapping Division
DNR- Maryland Department of Natural Resources
UCSB- University of California at Santa Barbara
Hunter- Hunter College

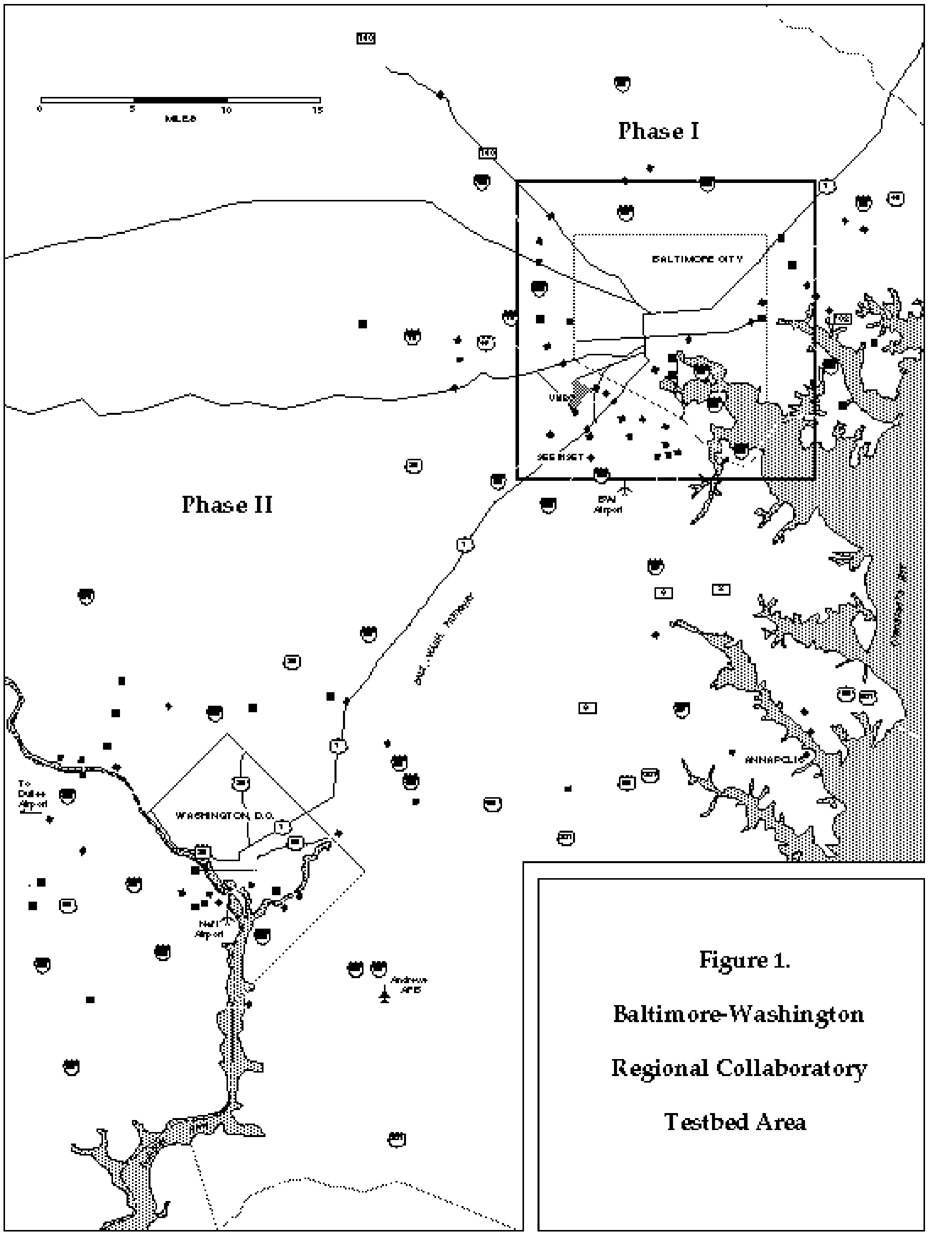


Figure 1.

Baltimore-Washington

Regional Collaboratory

Testbed Area

This multi-year collaboration continues to support activities assembling an integrated and flexible temporal urban land characteristics database encompassing the period from 1792 to 1992. Phase I focused on testing the HILT methodologies for an area encompassing greater Baltimore metropolitan area. Database design and construction, metadata documentation, and basic visualization methods have been tested and implemented using the Phase 1 database (Acevedo et al. 1996, Masuoka et al. 1995). Phase II efforts are currently expanding the database development for the entire 2-degree by 2-degree region. Phase III and IV will focus on experimenting with selected mapping themes, analyzing spatial patterns and rates, and linking with various environmental models. Included in the Phase I and II database development are temporal mapping layers for primary transportation, hydrography, and population density. Derivation of these data layers comes from the archives of historic maps and records prior to the 1970's and digital data in the post 1970's era using Landsat imagery, Digital Line Graphs, Digital Elevation Models, and DIME and TIGER files.

The expansion of activities and contributing agencies related to the Baltimore-Washington regional spatial database are lending support to the testbed concept, Figure 2.

Figure 2. Collaboratory Participants

In as much as this testbed provides a distributed, Internet accessible resource for environmental scientists as well as interested local and regional planners, the term "collaboratory" has been applied. Other activities associated with the Baltimore-Washington Collaboratory include data visualization research, NSDI metadata compliance and testing, user community outreach, and applied science modeling. A multi-disciplinary team has expanded on methodology, definitions, and collection criteria used to define the various data layers, ensuring consistency in data definitions and data collection techniques among the different collaborators.

BALTIMORE-WASHINGTON COLLABORATORY DESIGN

In the design of large spatial databases a variety of concerns and constraints must be taken into account to avoid outright failure or at a minimum reduce the inefficiencies of the system using the database structure (Marble 1988). Databases designers have adopted various approaches to addressing the creation of spatial databases, and databases in general, which almost universally begin with defining the user requirements or functional requirements of an individual, organization, or application process or model. From this setting the designer can construct the bounding parameters of systems, data, people, and financial resources into a conceptual template for the database to exist and perform its primary function. Systems designers have many models to follow but the similarity among these models is more distinctive than the minor, and more often, semantic differences (Pressman 1987, Teorey and Fry 1982, King 1984). In GIS applications, approaches to successful systems designs were adapted from the more ubiquitous non-spatial systems engineering protocols (Calkins 1982, Calkins and Marble 1987). Today, GIS consultants use essentially the same design to develop an integrated GIS installation and the requisite database for businesses or municipalities. Experience with "successful" installations has kept the system design process more narrowly defined than many outsiders might expect, to the point that "boilerplate" assessments comprise a significant portion of many contracted design studies. This does not demean the professionalism of GIS design consultants, who must have the expertise to know what is boilerplate and what is not, but illustrates that the process of system design and database design follows a structured series of fundamental steps from design to implementation.

For the Baltimore-Washington Collaboratory, the design of the spatial database follows a new paradigm outside the traditional GIS construction domain. This new paradigm is based on the realities of the collaborating data providers and users and therefore can be described from a couple of perspectives. From the data providers perspective the Collaboratory is following the general constructs of both the National Spatial Data Infrastructure (NSDI) and the NASA Mission to Planet Earth (MTPE) plans for handling Earth observations from space (National Research Council 1994, NSTC 1995). From the data user community another set of perspectives is evident which includes environmental modelers who will plan to use the Collaboratory assets as not only model input but as a means to calibrate and validate their models (Oreskes et al. 1994, FGDC 1995). Other users will likely view the Collaboratory as a source of input for regional planning purposes to analyze and predict rates of land use change and establish the causal factors related to the land use transitional processes. The most challenging user community represents those developing integrated regional models to couple environmental, human, and physical models (Blood 1994, NSTC 1995). This latter group of modelers will be instrumental in redefining both the identity of content and structure of the Collaboratory's assets. It therefore becomes incumbent on the Collaboratory designers and modeling community to creatively deal with issues of calibration, validation, uncertainty and error propagation, simplification or aggregation, resolution and scale as they impact the performance of integrated regional models or environmental models in general. These issues serve as primary assumptions in the Collaboratory design and therefore must in part be assessed on a case by case basis with individual modeling teams but with attention to the ramifications of any design constraints on the use of the spatial database for general modeling applications. With these caveats, a conservative approach has been applied to the creation and documentation of the spatial database.

Using contemporary GIS capabilities, the Baltimore-Washington Collaboratory assumes data assets to be either digital vector or raster with associated attribute files and metadata. Data sets initially represented in the Collaboratory, Figure 3 encompass:

- multi-temporal,
- multi-scale,
- multi-resolution,
- multi-spectral, and
- multi-thematic data resources.

It is assumed that most vector datasets will be converted to grid formats for input in cellular autonoma, finite element or finite difference modeling structures (Coucleilis 1985, Clarke et al. 1996). Other uses of the vector data sets would include referencing of geographic phenomena, via hydrology or transportation alignments or as vector overlays for improved comprehension of associated datasets. Error propagation attendant to vector-to-raster conversion remains unavoidable and will necessarily be the responsibility of the user (Lunetta et al. 1991).

Digital data resources of the Baltimore-Washington Collaboratory will be accessible via the Internet and in bulk media formats. Physical location of the Collaboratory's spatial database assets is distributed among the cooperating data providers with a few nodes accepting additional responsibilities to serve as Regional Data Centers (RDCs) under the guidance of NASA's Earth Observing System Data and Information Systems (EOSDIS) general protocols. A significant assumption for the Collaboratory is that data providers (e.g., NASA, USGS, NOAA) should remain stewards of data generated whenever possible to keep data redundancy to a minimum, maintain metadata documentation, and provide updates and upgrades to data as appropriate. The Collaboratory will ensure that data resources included in the data catalog have been compiled with cooperative protocols as defined later in the text under Documentation/Metadata. The RDC design acknowledges that many of the digital data resources will require various preprocessing steps to make the data suitable for some applications for local or regional users and environmental modelers. For example, AVHRR or Landsat data may require some data format handling before use on PC based GIS software packages by county planners and decision makers.

In functioning as a Regional Data Center, the Baltimore-Washington Collaboratory, using personnel from UMBC and NASA Goddard Space Flight Center are working with various environmental modelers. One example of an environmental application incorporates the testing of the Hydrologic Simulation Program in Fortran (HSPF) hydrologic model for performance along an urban-rural gradient using input parameters from the Collaboratory resources. The modeling evaluation, in cooperation with personnel of the USGS, Yale School of Forestry, and the Institute of Ecosystem Studies, entails examining how HSPF performs along this gradient at cascading spatial scales or grid resolutions. The RDC functions for the Collaboratory then serve in an iterative fashion to both supply data for modeling and to share the results back to the community for assessment and planning purposes. Another modeling application entails providing data for testing and calibrating an urban growth cellular autonomon model developed by Clarke (1996). The results of this model are planned to be available over the Internet, including operating code and data. This approach will provide local land use managers with virtually no-cost tools to examine population growth for their regions 50 years into the future, while also allowing them to analyze the past growth phenomena with the Collaboratorys historic land use and demographic data sets.

The importance of using the Baltimore-Washington Collaboratory to verify, calibrate, and validate researcher's models includes research on performance of remotely sensed data. By attending to careful geographic registration of the regional infrastructure, land use, demographics, topography, and other physical data sets as part of the Collaboratory shared digital resources, remote sensing scientists can utilize Collaboratory resources for ground truth calibration. This will become increasingly important as a host of new sensors being developed for the EOS program begins to produce data. In addition, the user community of local environmental and land use managers and planners and commercial entities will look to the RDC/ Collaboratory data sets to determine the applicability of EOS information for their local/regional applications.

COLLABORATORY DOCUMENTATION/METADATA

A variety of data integration processes are involved in the creation of the Baltimore-Washington Collaboratory assets that must be explicitly defined for the user community. These definitions are included under the metadata design protocols directly from the federal metadata standards (FGDC 1994). Both federal and Maryland state agencies have been directed to comply with these standards. While the data provided to the Collaboratory varies in format and quality, the attention to the details of metadata documentation provide environmental modelers with the information required to determine goodness-of-fit for their modeling use. While the federal metadata standards have been viewed as unfunded mandates, compliance provides the modeling community with a rich resource of digital data that would otherwise be risky in terms of adding uncertainty to their modeling parameters (Berk 1994). This adherence to metadata documentation is not a trivial exercise and has required significant use of project personnel resources but serves as critical input into the NSDI national resources. Numerous technical problems have been discovered in the implementation of the FGDC metadata standards, solutions to these implementation problems should help with the design of future NASA RDCs as NASA will need to understand the requisite administrative overhead necessary to keep in FGDC compliance.

Initial results from the Baltimore-Washington Collaboratory have demonstrated the usefulness of an hierarchical approach to metadata documentation under a hybrid FGDC schema currently under testing. The hybrid approach is designed to streamline the inclusion of local and regional digital data resources from agencies that do not comply with federal or state standards. This will enable an increase in data resources available through the Collaboratory at fine resolution scales (1 meter to 10 meters) while still attending to the philosophy for goodness-of-fit labeling requirements, Figure 3. Performance testing of the hybrid schema is scheduled for summer of 1996.

Figure 3. Baltimore - Washington Regional Collaboratory Baseline Data Resources
Proposed Thematic/Spatial Coverage

Major Data Base Thematic Categories	Map Scale/Information Level or Source			
	Local/County	Regional		Global
		State	Federal	
• Land use/land cover -wetlands -soils -ecosystems -hydrology -topology -hazards/toxics	(1 meter - 0.5 hectare) • Digital orthoquads • Historic aerial photos • Digital records	(5 hectare) • Thematic mapper • MD landbase • GAP	(15 hectare) • TM • GAP • DEM • DLG • AVHRR, EOS prototype, NASA experimental sensors, radar, thermal, et cetera	(100 hectare)
• Population/ demographics	•Subdivision plats	• Assessment and Taxation Office	•Census -STF 1A -STF 1B	• DMSP/OLS
• Transportation	• Street network	• MD landbase	• DLG	
Potential Datasets				
• Economics	• Property tax records	• MD Office of Planning	• Dept. of Commerce • Census	• World Bank • NRC
• Energy	• Municipal digital facility infastructure records	• MD Office of Planning	• DOE	• World Bank • NRC

CONCLUSION

The creation and utilization of a regional spatial data testbed is critically needed in the environmental community to calibrate, verify, and validate the various models. This requires that the design of a regional, digital spatial data resource be established in a general GIS structure to support the ready import of data into a variety of environmental models. Development of spatial models continues to mature for multi-scale, multi-temporal, and multi-thematic applications under a variety of schema for entity-attributes. Progress has been reported by Raper and Livingstone (1995) and others (Guptill 1988, Shi and Zhang 1995) in the use of object or feature base representation of spatial, temporal, and attribute modeling. A regional digital database must attend however to the needs of the many and therefore a conservative GIS structure was selected for the Baltimore-Washington Collaboratory. In addition, by aligning the Collaboratory protocols with national standards and trends, modelers using the digital data resources can benefit from working in the context of the national spatial data infrastructure where efforts to correct and improve the national protocols will result in more meaningful approaches for long-term use of their models.

Environmental modeling needs to be understood from a broader context than the disciplines of origin. Integrated assessment models and integrated regional models will require increased understanding of the semantics and parameter formats of different modeling schools. Environmental models using GIS data structures and resources will require extension into the domains of human ecology, urban environments, landscape ecology, sustainability, and ecological economics to meet the demands for improved decision making and management applications. It is envisioned that through the application of quality documented data resources, available from RDCs such as the Baltimore-Washington Collaboratory, development of integrated environmental modeling can be better accomplished in the future.

ACKNOWLEDGEMENTS

We wish to thank Janet Crawford and Susan Clark of the USGS in Reston, Virginia for their assistance in this project. The work was funded in part by NASA Research Grant NAGW-1743.

REFERENCES

Acevedo, W. and C. Bell. (1994) Time series animation of historical urban growth for the San Francisco Bay region. Abstracts Association of American Geographers 90th Annual Meeting. San Francisco, CA, pp. 2.

Alig, R.J. and R.G. Healy. (1987) Urban and built-up land area changes in the United States: an empirical investigation of determinants. Land Economics. Vol. 63, pp. 215-226.

Asrar, G. and J. Dozier. (1994) EOS: Science Strategy for the Earth Observing System. AIP Press, Woodbury, NY.

Bell, C., W. Acevedo and J. T. Buchanan. (1995) Dynamic mapping of urban regions: growth of the San Francisco Sacramento region. Proceedings, Urban and Regional Information Systems Association. San Antonio, TX, pp. 723-734.

Berk, Richard A. (1994) Uncertainty in the construction of interpretation of mesoscale models of physical and biological processes. In P. Groffman and G. Likens (eds.). Integrated Regional Models: Interactions Between Humans and Their Environment, Chapman and Hall, New York, NY, pp. 50-64.

Blood, E. (1994) Prospects for the development of integrated regional models. In P.M. Groffman and G.E. Lines (eds.). Integrated Regional Models: Interactions Between Humans and Their Environment. Chapman and Hall, New York, NY, pp. 145-153.

Calkins, H. W. (1982) A pragmatic approach to geographic information system design. In Pequet and O'Callaghan (eds.). Proceedings U.S./Australia Workshop on Design and Implementation of Computer-Based Geographic Information Systems. Amherst, NY: IGU Commission on Geographical Data Sensing and Processing.

Calkins, H. W. and D. F. Marble. (1987) The transition to automated, production cartography: design of the master cartographic database. The American Cartographer. 14 (2).

Clarke, K.C., J. A. Brass, and P. Riggan. (1995) A cellular automaton model of wildfire propagation and extinction. Photogrammetric Engineering and Remote Sensing, Vol. 60, No. 11, pp. 1355-1367.

Clarke, K., S. Hoppen, and L. Gaydos. (1996 forthcoming) Methods and techniques for rigorous calibration of a cellular automaton model of urban growth. Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21-25, 1996.

Coucleilis, H. (1985) Cellular worlds: a framework for modeling micro-macro dynamics. Environment and Planning , Vol. 17, pp 585-596.

FGDC. (1994) Content standards for digital spatial metadata. Federal Geographic Data Committee, Washington, D.C.

FGDC. (1995) National geodata forum. Federal Geographic Data Committee. May 1995, Reston, VA.

Farmer, D.G. and M. J. Rycroft. (1991) Computer Modeling in the Environmental Sciences. Oxford: Clarendon Press.

Guptill, Stephen C. and Robin G. Fegeas. (1988) Feature based spatial data models--the choice for global databases in the 1990's? In H. Mounsey and R. Tomlinson (eds.). Building Databases for Global Science, Philadelphia, PA: Taylor & Francis, pp. 279-295.

Kemp, K. (1993) Environmental Modeling with GIS: A Strategy for Dealing with Spatial Continuity. Technical Report 93-3, National Center for Geographic Information and Analysis, Santa Barbara, USA.

Kemp, Z. and A. Kowalczyk. (1994) Incorporating the temporal dimension into a GIS. In M. Worboys (ed.). Innovations in GIS. London: Taylor & Francis, pp. 89-103.

King, D. (1984) Current Practices in Software Development: A Guide to Successful Systems. New York: Yourdon Press.

Kirkland, D., L. Gaydos, K. Clarke, L. DeCola, W. Acevedo and C. Bell. (1994) An analysis of human-induced land transformation in the San Francisco Bay/Sacramento area. World Resource Review, Vol. 6, No. 2, pp 206-217.

Lunetta, R. S., R. G. Congalton, L. K. Fenstermaker, J. R. Jensen, K. C. McGwire, and L. R. Tinney. (1991) Remote sensing and geographic information system data integration: error sources and research issues. Photogrammetric Engineering and Remote Sensing, 57(6), pp. 676-687.

Marble, D. F. (1988) Approaches to the efficient design of spatial databases at a global scale. In H. Mounsey and R. Tomlinson (eds.). Building Databases for Global Science, Philadelphia, PA: Taylor & Francis, pp. 49-65.

Masuoka, P., T. Foresman, S. Fifer, W. Acevedo, S. Clark, J. Crawford, and J. Buchanan. (1995) Visualization techniques for the analysis of Baltimore regional GIS data. Proceedings Volume 2, GIS/LIS '95. Nashville, TN, pp. 704-712.

McDonnell, M.J. and S.T.A. Pickett. (1990) Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology. Ecology, Vol. 71, pp. 1232-1237.

Mitasova, H., L. Mitas, and W. M. Brown, D. P. Gerdes, I. Kosinovsky, and T. Baker. (1995) Modelling spatially and temporally distributed phenomena: new methods and tools for GRASS GIS. International Journal of Geographical Information Systems, 9(4), pp. 433-446.

National Research Council. (1994) Science Priorities for the Human Dimensions of Global Change. National Academy Press, Washington, D.C.

National Science and Technology Council (NSTC). (1995) Our changing planet: the FY 1995 U.S. global change research program. A Report by the Subcommittee on Global Change Research, Committee on Environment and Natural Resources Research of the National Science Technology Council.

Oreskes, N., K. Shrader-Frechette, and K. Belitz. (1994) Verification, validation, and confirmation of numerical models in the earth sciences. Science, Vol. 263, pp. 641-646.

Pickett, S.T.A. and M.J.Cadenasso. (1995) Landscape ecology: spatial heterogeneity in ecosystems. Science, Vol. 269, pp 331-334.

Peuquet, Donna J. (1994) It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems. Annals of The Association of American Geographers, Vol. 84, No. 3, pp. 441-461.

Pouyat, R.V. and M.J. McDonnell. (1991) Heavy metal accumulations in forest soils along an urban-rural gradient in southeastern New York, USA. Water, Air, and Soil Pollution. Vol. 57-58, pp. 797-807.

Pouyat, R.V., M.J. McDonnell, and S.T.A. Pickett. (1995) Soil characteristics of oak stands along an urban-rural land-use gradient. J. Environmental Quality, Vol. 24, pp. 516-526.

Pouyat, R.V., R.W. Parmelee, and M.M. Carreiro. (1994) Environmental effects of forest soil-invertebrate and fungal densities in oak stands along an urban-rural land use gradient. Pedobiologia., Vol. 38, pp. 385-399.

Pressman, R. S. (1987) Software Engineering: A Practitioner's Approach. (Second Edition). New York: McGraw-Hill Book Company.

Raper, J. and D. Livingstone. (1995) Development of a geomorphological spatial model using object-oriented design. International Journal of Geographical Information Systems, 9(4), pp. 359-383.

Shi, W. and M. Zhang. (1995) Object-oriented approach for spatial, temporal, and attribute data modeling. Proceedings Volume 2, GIS/LIS '95, Nashville, TN, pp. 903-912.

Teorey, T. J., D. Yang, and J. P. Fry. (1986) A logical design methodology for relational databases using the extended entity-relationship model. Computing Surveys, 18(2), pp. 197-222.

Vitousek, P. (1994) Beyond global warming: ecology and global change. Ecology, Vol. 75, No. 7, pp. 1861-1876.

AUTHOR INFORMATION

Timothy W. Foresman
Director, Spatial Analysis Laboratory
Department of Geography
University of Maryland Baltimore County
5401 Wilkens Avenue
Baltimore, Maryland 21228
Phone: 410-455-3149
Fax: 410-455-1056
Email: foresman@umbc.edu

Loey Knapp, Patricia Andrews, and John Turek

Image Navigation for Wildland Fire Location Mapping

Paper for the NCGIA Conference on GIS and Environmental Modeling, January 21-24, 1996, Santa Fe, New Mexico.

ABSTRACT

Efficient response to wildland fires is of critical concern to various state and federal agencies, and requires a substantial data archive from which to draw essential information; examples are maps of existing regional fires, fuel type or vegetative greenness, and populated areas that might be threatened. For effective response to the fire, these data must be up to date and easily accessible for prediction, planning, and resource allocation but typically this is not the case. While spacecraft imagery has long held the promise of global data acquisition and real time spatial information, the analysis process has proved to be a bottleneck due to the vast amounts of data and the size of the datasets. The challenge in the area of emergency response is to provide the technology to extract crucial information from images quickly enough to influence the decision-making process.

This paper will describe new technology in the area of content-based search of images which can be applied to emergency preparedness and emergency response. This technology will be discussed in light of a specific emergency, that of wildland fire.

INTRODUCTION

Wildland fire management in the United States is the responsibility of various agencies ranging from federal and state to rural and private. Fire, however, doesn't recognize administrative boundaries. This has led to interagency and international cooperation in detection and response to wildfires. In some cases, dispatchers and coordinators from various agencies are co-located to facilitate cooperation and information sharing. Nevertheless there is room for improvement in tracking fire activity on a national or regional level across all land ownerships as well as in archiving historical fire data. National or regional mapping of fire locations from satellite data would be useful additional information if it were easily accessible across multiple agencies and up to date; if it can be combined with other ancillary data such as topography, urban areas, and administrative boundaries; and if it can be used in conjunction with fire potential models.

Efforts to provide regional and/or national information on fires have taken different forms. The first, official collection and distribution mechanism is through the National Interagency Fire Center (NIFC) in Boise, Idaho. Local agency reports are fed to NIFC for consolidation, analysis, and reporting purposes. Daily fire situation reports from NIFC include a narrative on fires and areas of interest plus summaries of number of fires and acres burned by agency and region of the U.S. and also for Canadian Provinces. (The reports can be found on the Forest Service home page, http://www.fs.fed.us under "Caring for the Land."). The reports are compiled from intelligence information collected from geographic area coordination centers and are dependent on agency-specific input and cooperation.

Another source of wildland fire information is satellite imagery, a source of data which crosses political boundaries. The National Geographic Data Center (NGDC) within the National Oceanographic and Atmospheric Administration (NOAA) has developed classification algorithms for fires using the Defense Mapping Satellite Program (DMSP) operational linescan (OLS) sensor. These algorithms are used on the visible band of the night orbit and are adjusted for clouds, city lights, and lightning (Elvidge and others, in press). This technology could result in an Internet product of fire location maps for the U.S. on a daily basis.

The use of satellite imagery provides national data and avoids problems of reporting delays and consolidation, but a daily map product also has its limitations. A map product is by definition a predefined display. As in any GIS work, the capability to interactively search the imagery in conjuntion with ancillary data sets, would provide the analyst with more information on spatial relationships. In the case of wildland fire, the search must include temporal functionality as time is an essential element of wildland fire analysis. The technology to search images by color, texture, and shape has been developed and is used in multimedia software. An extension of this concept can be used to search multispectral images based on classification algorithms, color, or spectral signature.

Using a wildland fire classification algorithm an analyst could search a sequence of images for current fire locations, examine the temporal extent of a fire, or compare different years/seasons. Including additional classification algorithms the analyst could pose additional queries to determine wildland fires near urban areas or to find other spatial areas exhibiting similar characteristics to those currently burning, e.g. very dry with heavy fuel load. The resulting map of fire locations, combined with ancillary spatial data sets, could provide supplemental information such as fire on all land ownerships or fires per land cover type. Use of satellite-based fire location maps, like the NIFC fire situation reports, would be in monitoring the broad view of fire activity, and not in the management of individual fires.

Provision of an image search capability for wildland fires via the Internet would address the interagency aspect of the wildland fire problem. The primary advantage of Internet over private networks is that it provides general accessibility to information. This accessibility makes Internet useful for individuals or groups which are organizationally unrelated but have a common concern relative to a particular issue, such as the USDA Forest Service, the Bureau of Land Management, and the State forestry department's common interest in wildland fires. Providing remotely-sensed wildland fire information via the internet would solve two problems; 1) inclusion of regional or national in addition to agency data, and 2) accessibility to that data to state and federal, or public and private, analysts.

The rest of this paper will discuss the potential application of image search in the area of wildland fire location mapping, the application-specific challenges involved in implementing such a system, and the technology challenges which must be overcome for Internet-based image search to meet the implementation needs. This discussion will reference a system currently being architected by IBM that will be tested in a wildland fire application as a USDA Forest Service research project. The purpose of the collaboration is to test new technology in an implementation scenario and, by doing so, determine operational roadblocks.

WILDLAND FIRE LOCATION MAPPING

Primary users of a broad area fire location map product are expected to be in interagency coordination centers. Image queries and the resulting displays can be used as a focus of discussion during morning briefings to agency administrators. The information can be used in assessing the overall fire situation as well as in setting priorities for distribution of limited resources including aircraft, crews, and supplies. Managers in regional centers can view the fire activity in their area of responsibility compared to other areas of the country. The maps and related information can also be provided to the media, the public, and special interest groups, aiding coordination center personnel in carrying out an important function during times of critical fire activity.

In addition to timely fire intelligence information, fire history data can also be queried. Satellite data could be used to determine correct lat/long locations for fires, overcoming errors that often occur in determining and recording fire location. Wildfire data are archived by each U.S. federal agency and by States in separate data bases. Satellite data, of course, can't provide information on cause, cost, etc. so it is not a subsititute for those data bases. But an interagency fire location data base derived from the satellite data would aid post season analysis and planning activities.

Fire location mapping on a broad area will be especially useful when it is combined with GIS ancillary data such as political boundaries (State, county, etc.), land ownership (Forest Service, National Park Service, State, etc.), cities and roads, lakes and rivers, lat/long grid, and vegetation and fuel type. Besides giving a reference for location of fire activity, this information can be used in data searches. A fire manager might want to identify and display fires on Bureau of Land Management land or those in wilderness areas, fires within a specified distance of major cities, or fires that are on forested land.

Links to fire models will add information on fire potential to that on current fire activity. Fire locations can be overlayed on data layers generated by Wildland Fire Assessment System (WFAS), the next generation U.S. fire danger/behavior system (Andrews and others, in press). Products that are currently available to fire managers include a weekly greenness map derived from AVHRR satellite data indicating the state of live vegetation (Burgan and Hartford 1993, Hartford and Burgan in press), a daily fire danger map derived from weather taken at fire weather stations through the U.S., and a daily Haines Index (stability and dryness) map derived from upper air soundings from all North American weather service stations. The capability to "zoom in" to access data on topography and fuel type, would allow a link to site-specific fire behavior models (Andrews 1986) and fire growth simulation models (Finney in press, Finney and Andrews in press).

APPLICATION CHALLENGES

Implementation of an Internet-based image search capability for wildland fires poses several challenges. The first of these is in the area of data. Various sensors can provide information on wildland fires but no single sensor has the spatial resolution, temporal resolution, and spectral range required to provide all the information required by fire analysts. It may be that a data archive specific to wildland fire search will need to include the output from multiple sensors and will also necessitate cross-sensor query capability.

Prevedel (1994) reported on the use of AVHRR satellite data to monitor wildfire activity in several western States for 45 days in 1994. Maps were provided to Multi-Agency Coordinating (MAC) groups that were formed to deal with the severe fire season. In addition to a description of the development and use of the map product, he pointed out some limitations of the satellite data. Sun reflection, lightning, and rocks heated by solar radiation, for example, can be difficult to distinguish from fires. The 1.1 kilometer resolution is also a drawback to identifying and locating small fires.

The NGDC project has made significant progress with the Defense Meteorological Satellite Program (DMSP) operational linescan (OLS) data which uses night-time visual sensors. This sensor eliminates some of the problems with the use of AVHRR, but suffers from some of its own. This data has a spatial resolution of .5 kilometer, an improvement over AVHRR, and comparative studies between the sensors indicate that more fires were located with DMSP OLS data (Elvidge and others, in press). However, fires in urban areas can be missed and thick smoke can keep the satellite from locating fires. Also, the use of the visible band from the night orbit limits the temporal resolution to daily observations.

The Geostationary Operational Environmental Satellite Visible Infrared Spin Scan Radiometer Atmospheric Sounder (GOES VAS) sensor, which provides measurements in the visible and infrared regions, is another potential fire data source. The temporal resolution of 30 minutes is attractive but the spatial resoluton of .9 kilometers in the visible region and 6.9-13.8 kilometers in the infrared region limits the sensors use to location of very large fires or the monitoring of diurnal cycles of fire activity.

Research on the use of satellites for detecting and locating wildland fires is ongoing. When better techniques have been developed, tested, and become available, they will be used. Investments in two-meter resolution sensors are likely to bring significant payoff in this area towards the end of the century.

The second challenge is in the area of data representation. Prevedel noted that distribution of raw uninterpreted satellite data can lead to misinterpretation of the fire situation. For example, if a pixel is saturated it will indicate the presence of fire throughout the pixel regardless of the true spatial extent of the fire, thus overestimating the area of the fire. On the other hand, the inability of this technology to detect some fires due to size, clouds, smoke or other sensor problems will lead to an underestimation of both number of fires and area burning. Finally, 'counting' fires may cause problems as more than one small fire might be alight within a pixel or a single fire may have a gap, visually indicating two fires instead of one.

Due to these problems it is essential to consider the representation of the information to fire analysts and response teams. For instance, glyphs might be used to indicate a fire location rather than number of pixels beyond the threshold value. Color selection and provision of contextual information such as political boundaries must also be considered. These factors indicate the need for visualization capabilities beyond predefined image display which allow the user to specify the nature of the data representation. Such features are common in desktop software but not yet provided on the Internet.

TECHNOLOGY CHALLENGES

From a technology perspective, the application outlined above raises serious challenges. An interactive Internet-based image search system comprises several key elements including user-specified queries and displays, image search through classification algorithms or user defined parameters such as color, and data retrieval (download). In the latter case it is important to be able to reconstruct the image in a lossless manner as many applications cannot afford the loss of data.

These system elements cannot be implemented without the following technical capabilities:

1. compression with associated decompression to a lossless image
2. a spatial query language
3. classification algorithms which can be applied to compressed data
4. simple GIS functionality including georegistration, resampling, and aggregation against compressed data,
5. visualization operations, also for use with compressed data.

System Structure

The overall organization of the system can be seen in Figure 1. (not yet available) Clients connect to servers using standard internet protocols (i.e., http). This approach takes advantage of and builds on the tremendous infrastructure that is in the process of being developed.

Client Side Interface

The internet provides a common interface for access to information through standard Web Browsers like Mosaic or Netscape. Due to the availability of this interface on multiple platforms it would be desirable to base a system approach entirely on standard Web browsers. Unfortunately, even though new capabilities are constantly being added (e.g., JAVA) it is unlikely that browsers will provide the basis required to fully support the image manipulation required by a generic emergency management system. In particular, visualization tools remain particularly weak. The system under development takes a two-tiered approach. Limited functionality is available through a standard Web Browser while users who need to run complex queries can download an enhanced browser.

The enhanced browser supports a language that allows visualization and parameter specification to be integrated with requests to the server. One of the goals of the language is to preserve the illusion of a single session even if the session spans multiple requests to the server. More importantly, the language allows the user to dynamically define semantics of the queries and tailor the system to the specific needs of the environment with minimal effort.

Server Organization

The system server is based on the use of a standard http daemon. The server initiates a session by invoking the query parser described along with the rest of the major system components below: \begin{itemize} \item {\em Query Parser}: Queries that come into the system need to provide a mechanism that can freely specify new models for extracting features and relate them back to content. The query parser offers a general purpose language based on image set manipulations that allows for this kind of interaction. \item {\em Data Manager}: The data manager provides support for all ancillary data and indices. The current version of the manager uses IBM's DB2 engine. We will be providing extensions to experiment with the effectiveness of different index structures. \item {\em Image Manager}: Satellite and GIS data typically come in raster format that can be viewed as an "image" or a 2-dimensional lattice. The Image manager stores N-dimensional lattice data in transformed/compressed format that can be used to facilitate image analysis and feature extraction for content based search. This is described in further detail below. \item {\em Image Processing Engine}: The image processing engine is used to provide the ability to filter, process and do feature extraction on images during the search and retrieval process. These operations need to be closely tied to the representation of the data on disk as described below. \item {\em Visualization Engine}: Although it is necessary to provide visualization tools on the client side, in some cases it will be more effective to provide the final product directly from the server side. The current version of the visualization engine uses IBM's Data Explorer product. \end{itemize}

Challenges

The management of the large and diverse kinds of data required by an emergency management system pose significant technical challenges. Some of these are outlined in the remainder of this section.

Image Compression for Data Storage

In order to fully understand the nature of the problem at hand consider the following example: Landsat Thematic Mapper scenes are available at a resolution of 30 meters per pixel containing 7 spectral bands. These data could be quite useful in the event of an emergency, aiding in the location of urban/wildlife interfaces, water sources, etc.

Full coverage of the United States with Landsat imagery would require on the order of

100GB. Unfortunately, these images usually have partial cloud cover and so several scenes from each location need to be stored in order to ensure that timely data are available on the area of interest. In the near future, new commercial satellites are expected to have a resolutions down in the 2 meter range thereby increasing the storage requirements by 2 orders of magnitude.

Other sources of data, such as the OLS sensor that is one possible source of fire data, have reduced spatial resolution but increased temporal resolution. Thus, even though the price of storage devices will continue to drop at a dramatic rate, the cost of storing satellite images in an operational system will continue to be the dominant cost of the system. Proper use of compression can dramatically reduce this cost.

Compression techniques can be either {\em lossless} or {\em lossy}. A lossless compression scheme is one which guarantees perfect reconstruction of all of the bits in the original dataset. A lossy compression scheme, on the other hand, does not reconstruct most images exactly but rather allows the loss of some information in order to achieve higher compression ratios. The assumption is that only "unimportant" information is lost. Because it is difficult to determine exactly what information will be relevant in the analysis stage, our compression techniques need to provide a progressive framework that allows users to select exactly the level of loss that they can tolerate. The system needs to provide for lossless retrieval of data. The compression mechanisms under development achieve lossless compression that is comparable to the best lossless compression schemes currently available for image products while still maintaining the ability to do progressive retrieval.

Analysis of Compressed Imagery

Although extraction of relevant features at the time of data ingest into a predefined schema is an important part of any information system, it is by no means sufficient. The needs of each region and office can be quite different. As an example, the algorithm that is used for identifying fires in OLS data relies on detecting sources of visible and near-infrared emission on the earth's surface in the absence of solar illumination (i.e., at night). While this algorithm can be readily applied in places such as Brazil and parts of the continental U.S., it can not be applied in places such as Northern Canada and Alaska.

In essence the feature extraction process is by its very nature lossy and cannot represent all of the content contained in the imagery products that feed into the analysis required by the different regions. Furthermore, the processing involved in deriving the attributes can be quite expensive and techniques to make this processing more efficient are required. Thus, although useful and necessary, the use of a predefined schema will be insufficient to adequately support the search mechanisms required by an emergency management system. It is necessary to provide the functionality that will allow the user to visualize, define and extract features dynamically thereby performing content-based search in real-time on the image products rather than assuming that all information will be available at the time of the emergency.

Although it would seem that the cost associated with this processing is higher when images are stored using compression, this need not be the case; by laying out the data in an appropriate fashion, the search process actually becomes more efficient and since it is not necessary to look at all the bits associated with the original image. The primary technical idea behind our project is that it is possible to increase the speed of searching through images stored in a digital library while simultaneously reducing the storage requirements.

Search, Manipulation, and Distribution via Internet

Another problem that arises is based on the limited bandwidth imposed by our transmission medium. In the foreseeable future it is expected that most offices will be connecting through telephone at relatively slow transmission rates. As a result, it is essential that only the relevant information be returned to the user. The system provides this functionality via the integration of several mechanisms including server based search and content extraction, caching on the client and compression.

As one example, when dealing with complex images, it is important to be able to {\em navigate} through the image. Figure 2 (not yet available) shows a portion of a Landsat TM image of the Indiana/Kentucky border. The figure shows a scenario in which a manager can start with a large area view of a region of interest: i.e, they start with a 128x128 pixel representation of a 2048x2048 portion of the image. Here, one can clearly see the major features of the landscape including the Ohio river but cannot identify other relevant features. At each step of zooming in the resolution is doubled, while the size of the browse product remains unchanged. As the resolution increases, it becomes easier to identify features such as mountains, cities and roads for managing site specific information. By maintaining small viewing windows, the user has managed to isolate the information of interest in an interactive fashion yet the cost of data transmission has been kept low. Even across telephone lines, each iteration of the search will take only a few seconds.

Heterogeneous data sets

One substantial difficulty is the need to deal with different data sources and types, for example, vector data for maps, raster data for digital elevation maps and satellite imagery. Even when the data sources are similar, for example all raster products, the management and mapping of different resolutions pose additional difficulties. The prototype in developement provides a framework that allows the integration of additional tools and the coordination of data migration between those tools.

Scalability

Scalability is a serious concern for a project of this type. It is likely that the amount of data required for an operational environment and the user load on the system will pose major challenges. As part of the technology that is being applied to the system server, IBM is exploring the use of parallelism and distribution to support larger user groups and the use of hierarchical storage formats to support large data volumes.

CONCLUSION

Numerous problems involve multiple agencies and require a centralized data archive specific to that problem but accessible by all of those concerned, public and private alike. One such problem is wildland fire, which is of interest to federal, state, and private land management groups. Other such problems are noxious weed infestation, agricultural plant health, and response to natural hazards. Within these problem areas, data collection across political boundaries and accessibility to the data can be major issues. The use of satellite imagery, accessed via the Internet, may address these problems.

In this paper, we have outlined the application of new technology in content-based search of satellite imagery over the Internet to the area of wildland fire. Significant application and technical challenges must be addressed for this technology to become operational. With successful resolution of these challenges, however, content-based search can be applied across a range of emergencies as well as to other pressing cross-agency problems.

ACKNOWLEDGEMENTS

This work is partially supported by NASA contract NCC5-101

REFERENCES

Andrews, Patricia L., (1986) BEHAVE: Fire Behavior Prediction And Fuel Modeling System-- BURN Subsystem, Part 1. General Technical Report INT-194. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 130 p.

Andrews, Patricia L., Bradshaw, Larry S., Burgan, Robert E., Chase, Carolyn H., and Hartford, Roberta A., (in press) WFAS: Wildland Fire Assessment System--Status 1995, Presented at 1995 Interior West Fire Council Meeting, St. George, Utah. Nov. 1-3, 1995.

Burgan, Robert E., and Hartford, Roberta A., (1993) Monitoring Vegetation Greenness With Satellite Data. Gen. Tech. Rep. INT-297. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 13 P.

Castelli, Vittorio, C.S> Li, I. Kontoyiannis, and J.J. Turek, Progressive Classification of Satellite Images in a Wavelet Framework, NJIT, September 22, 1995.

Elvidge, Christopher D., Herbert W. Kroehl, Eric A. Kihn, Kimberley E. Baugh, Ethan R. Davis, and Wei Min Hao, (in press) Algorithm for the Retrieval of Fire Pixels from DMSP Operational Linescan System Data, "Global Biomass Burning", edited by Joel S. Levine.

Finney, Mark A. (in press) FARSITE, A Fire Area Simulator for Fire Managers. Presented at The Biswell Symposium, February 15-17, 1994, Walnut Creek CA.

Finney, Mark A., and Andrews, Patricia L., (in press) The FARSITE fire area simulator: Fire management applications and lessons of summer 1994. Presented at 1994 Interior West Fire

Council Meeting and Symposium, Coeur d'Alene, ID, November 1-3, 1994.

Hartford, Roberta A., and Burgan, Robert E., (in press) Vegetation Condition and Fire Occurrence: A Remote Sensing Connection. A Paper presented at the Interior West Fire Council Meeting and Symposium, Coeur d'Alene, ID, November 1-3, 1994.

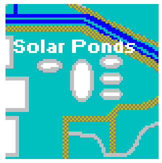
Prevedel, David A. (1994) Project Sparkey: A Strategic Wildfire Monitoring Package Using AVHRR Satellites and GIS Photogrametric Engineering and Remote Sensing 60:1 271-278.

Loey Knapp, Senior Analyst
IBM Corporation
6300 Diagonal Highway
Boulder, CO 80301
phone: 303-924-0482
fax: 303-924-0518
email: loey@vnet.ibm.com

Patricia Andrews, Project Leader
USDA Forest Service
Intermountain Research Station
P.O. Box 8089
Missoula, MT 59807
phone: 406-329-4827
fax: 406-329-4825

John Turek, Project Leader
IBM T.J. Watson Research Lab
P.O. Box 704
Yorktown Heights, N.Y. 10598
phone: 914-784-7591
fax: 914-784-6031
email: jjt@watson.ibm.com

David N. Fogel



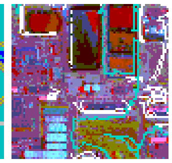


Image Rectification with Radial Basis Functions: Application to RS/GIS Data Integration

---

Back to top

Abstract

Remotely sensed digital imagery is inherently in need of geometric correction to maximize its usefulness within a GIS for subsequent analyses. This has long been problematic for large scale imagery such as acquired from airborne scanners. Radial basis functions, such as Hardy's multiquadrics and reciprocal multiquadrics, thin plate splines and variations of these methods offer ready alternatives to conventional polynomials and finite element methods for correction of these data. The nonstochastic methods of multiquadrics, thin plate splines and polynomials are reviewed in the context of scattered data interpolation and approximation. Applications to actual remotely sensed data acquired from airborne and satellite remote sensing systems (passive and active sensors) indicate that multiquadrics, reciprocal multiquadrics and thin plate splines produce excellent results while maintaining control point correspondence. Implementation specifics of these analytical functions, e.g. the development of the multiquadric

algorithm for image warping, are reviewed based on the development of the software program MQReg. MQReg was conceived and implemented as a research tool during NCGIA Initiative 12: Integration of Remote Sensing and GIS. The status of MQReg as a research tool, its availability and extensions to current research are also discussed.

Back to top

Introduction

Remotely sensed data preprocessing generally refers to geometric and radiometric corrections by data providers, to varying degrees, as well as subsequent image registration, rectification or orthorectification and further radiometric correction. Preprocessing by data providers (e.g. SPOT, EOSAT, USGS) infers a sense of data fidelity to the user who may not in fact be aware of the nature and the extent of the operations on the "raw" data. For example, the Landsat-5 imagery in Figures 1 and 2 illustrate different degrees of data preprocessing (Wivell 1995). However, far less stringent requirements apply to data fidelity with respect to subsequent geometric "preprocessing" by the users in applications such as change detection and classification. Specifically, this paper addresses this subsequent geometric preprocessing by reviewing conventional practice, offering new directions and providing an illustrative examples of the potential in this areas as an impetus for change. In comparison, radiometric preprocessing is a rather more dynamic research area, however it is beyond the scope of this paper.

Integrating remotely sensed data into a geographic information system provides valuable input for environmental modeling much the same way ancillary data has been used within digital image processing systems for classification and change detection. Common to both processing flows is remotely sensed data preprocessing to a format useful for subsequent analyses. This preprocessing requirement is increasingly important given developments in sensor technology and the increasing data stream of high resolution remotely sensed data. These new data are being acquired from both satellite and airborne sensors.

Issues surrounding the integration of remotely sensed data into a geographic information system (GIS) warranted the establishment of Initiative-12 by the National Center for Geographic Information and Analysis (NCGIA) in response to interests in both the public and private sectors. Information about the Initiative research agenda and substantive issues may be found in the proceedings from a special session at the 1991 ACSM-ASPRS Annual Convention (Star 1991) and NCGIA Technical Paper 93-4 (Estes and Star 1993).

Figure 1: Landsat-5 Image with Level 0R Correction.

Remote Sensing/Geographic Information System (RS/GIS) data integration is not a new topic. A raster-based information system for georeferenced data was coupled to an image processing system developed at the Jet Propulsion Laboratory (JPL) at the California Institute of Technology. This system is often referred to colloquially as VICAR/IBIS (Video Information Communication and Retrieval/Image Based Information System). Conceived as a subset of VICAR, IBIS included the ability to manipulate vector data. Marble and Peuquet (1983) provide an overview of GIS, RS/GIS, and an application using VICAR/IBIS.

Figure 2: Landsat-5 Image with Level 1G Correction.

This paper will review digital image warping, i.e. image registration and image rectification, in the context of change detection. Trotter (1991) provides an overview of the use of remotely-sensed data as an input to a GIS. Brown (1992) provides a good useful overview on image registration in general. Warping, as used here, refers to geometric transformations of digital imagery. An assessment of image registration issues and requirements follows in the form of a critique of conventional practice and a review of scattered data interpolation methods within a RS/GIS framework. Radial basis function methods are introduced and applied to airborne scanner data at the Department of Energy's Rocky Flats facility in Colorado.

Back to top

Background

Three broad categories of geocorrection techniques may be identified for post-preprocessing of remotely sensed data based on inputs to the procedure: (1) correction models based on ephemeris platform and sensor information (scene/sensor models); (2) correction models relating corresponding ground control points or features; and (3) hybrid models. In (2) above, ground control may include vertical (height) information as well as relative location and features may be extended to include collections of points. The focus here is on the second category, or more specifically, on mathematically modeling the geometric distortions based on control point correspondence. This research is biased toward airborne scanner data which is intrinsically more difficult than satellite data to geometrically correct. It is assumed that, in some sense, the imagery is reasonable and not given to excessive sensor or platform variations.

Bivariate mapping polynomials are the most commonly used method for the registration and rectification of digital imagery. An early application of bivariate mapping polynomials to image warping is documented in Markarian, Bernstein et al. (1973) for the correction of "high-resolution" images from the ERTS series RBV (Landsat Remote-Beam Vidicon sensor system). Computing resources were comparatively scarce and expensive and these polynomials of the trend surface form are simple and can be reasonably effective. Regardless, these bivariate polynomials were still too computationally intensive for all subsequent preprocessing tasks.

One solution to this problem involved generating a grid from the ground control points based on an underlying geometric model (e.g. bivariate polynomials, piecewise linear methods) and approximating the warp. Thormodsgard (1987) provides information on this gridding approach which offers a trade-off between geometric fidelity and computational intensiveness. With satellite data, these methods are widely applied and most often used when registration or rectification is required. Grid density may be adjusted depending on the nature of the geometric distortions such that the amount of information lost during the gridding approximation is minimized.

The use of bivariate polynomial mapping functions in remote sensing is the most widely known approach to image warping despite its well-known shortcomings. Christensen, Jensen et al. (1988) describe common frustrations using bivariate mapping polynomials to warp airborne scanner data (5.6 meters-squared). In this case, geometric correction is both labor and computationally intensive. The lack of attention to geometric correction is evident in publications where additional geometric preprocessing has taken place with no mention of the method (e.g. Curran and Pedley 1990). It must be conjectured that bivariate polynomial mapping functions were applied to warp the data since the approach is most prevalent and rarely contrasted against alternatives.

Seriousness consequences of misregistration are underscored in the change detection literature. In a review of change detection methodology, Singh (p. 990, 1989) states, "It may be mentioned here that accurate spatial registration of the two images is essential for most change detection methods." On a global scale, Townshend, Justice et al. (p. 1054, 1992) conclude that more comprehensive satellite preprocessing is required for change detection stating, "The registration of data sets to a common spatial framework is an essential precursor to the use of remotely sensed data for monitoring change." Townshend et al. conduct their research using third-order bivariate polynomials seemingly because "... geometric correction is [polynomials are] ... an established technique." (Heard, Mather and Higgins 1992).

A well-reasoned statement by Billingsley (p. 429, 1982) provides some perspective, "Given a certain loss in accuracy due to misregistration, how does that damage the ability to use the data analysis results? These evaluations will be discipline dependent, and must be sought separately." Billingsley's observation does not permit geometric distortions to be assumed away (out of convenience). It is of little consequence that there exists the common practice of reporting error figures loosely summarizing the warping accuracy. These root mean squared error (RMSE) figures usually underestimate misregistration. Interestingly, proprietary systems may not even calculate the same values for a metric as benign as a RMSE (cf. ERDAS Imagine 8.1 and PCI EASI/PACE 5.3).

Back to top

RS/GIS Data Integration

A basic knowledge of digital image processing and remote sensing fundamentals is requisite to understanding RS/GIS data integration issues. Reviews of remote sensing concepts may be found in Duggin and Robinove (1990) and Forshaw, Haskell et al. (1983). A digital image is basically a two-dimensional array of pixels having both radiometric characteristics and spatial components. The radiometric resolution is a function of spectral wavelength and bandwidth (among other things) and operationalized as a digital number (DN). Spatial resolution refers to a measurement or observation's local nature most often represented as a raster data type with relative location and size attributes. These attributes are not always sufficient to define spatial resolution since spatial resolution is not necessarily equivalent to pixel size (Forshaw et al. 1983). Remotely sensed data are often resampled to an arbitrary or convenient size bearing little relationship to the original spatial resolution.

Geometric distortions are generally classified as either internal sensor-related effects and external effects (Bernstein 1983). We will assume that most of the internal effects have been corrected in a prior preprocessing operation by the data provider (e.g. EOSAT). The principal distortions remaining are limited to internal effects (low-frequency) as well as high-frequency attitude and altitude effects (sensor-platform variations), and terrain effects (horizontal displacements). Airborne scanner distortions may be complicated by pronounced panoramic effects as well as S-Bend scanning effects.

A clear distinction between image registration and image rectification may be made based on whether the output domain is an image, a georeferenced or geocoded image, or a map (projection). In the most basic sense, registration refers to an image-to-image geometric transformation. Rectification may be either an image-to-geocoded image transformation or an image-to-map transformation. Significantly, there is not explicit modeling and removal of terrain effects. Orthorectification involves image rectification with explicit terrain effects removal. The distinction between registration and rectification is important with respect to discussing the ground control points (GCP) accuracy.

Image registration is perhaps the most straightforward procedure. It is likely that sufficient corresponding points can be identified given images with similar spatial and spectral characteristics. Extracting GCPs becomes more difficult when registering imagery from different data sources and/or at different resolutions. Even with proper attention to detail, it is reasonable to assume that there will be some error in both sets of GCPs, input and output (see, for example, the Total Least Squares Problem in Van Huffel and Zha 1993).

Image rectification is often more complicated. Recognizable features on a map are not always found on an image or may be difficult to recognize and vice-versa. Additionally, existing map accuracy standards may inhibit selection of representative GCPs. This situation arises when points in map space are less certain than the corresponding image points. This will become more problematic with the coming generation of high spatial resolution satellites.

In change detection applications, image-to-image registration or image-to-image rectification may have less geometric error than rectifying multiple images to a common map projection (see Niblack, 1986). On a case-specific basis, it may be determined that only the output product(s) be integrated into a GIS. If the output image is not orthorectified, then geometric distortions between images may be compounded during the correction process.

Digital orthophotos are not always well-suited for GCP extraction either. Clearly identifiable corresponding points are often lacking (e.g. cultural features) especially in the case of natural resource management. Rectification suffers from the same problems inherent in registration where errors exist in both the input and output GCPs. It is usually the case that these positional errors, horizontal and vertical, are ignored. This is true of digital elevation modeling (DEM) modeling as well where errors exist not only in the elevation data but in the locations of the elevation data too.

Back to top

Scattered Data Modeling

Geometric correction of imagery may be approached using bivariate functions where no assumptions are made with respect to the distribution of the locations of the control points in either map or image space. In fact, bivariate mapping polynomials are but one such method under this definition. Most literature on scattered data is in reference to interpolation. An interpolating function maintains the correspondence of the control points between images, whereas an approximating function will not honor the control point locations. Accordingly, it may be argued that the warping problem may be better approached as an approximation problem given the previously discussed difficulties in accurately locating corresponding GCPs. Bivariate mapping polynomials are most often approximating functions, though this is not justification for their (seemingly) universal application.

The expression "scattered data interpolation" was is used to distinguish applicable methods from otherwise structured-dependent techniques. Often, the objective is to obtain a contoured representation of the data. Survey literature may be found in reports, conference proceedings and monographs, as well as journal articles. These include, but are not limited to Franke (1982, 1987) Franke and Nielson (1991), Lam (1983), Lancaster and Salkauskas (1986), Watson (1992) and for 3D problems, Alfeld (1989). The RS/GIS community will recognize the potential applicability of contouring methods, terrain modeling schemes and data visualization techniques to image warping since these too are interpolation/approximation problems.

A distinction may be made between local and global models based on the whether or not the function varies with respect to location, e.g. piecewise linear functions vis-à-vis bivariate polynomials. The former involves a tessellation of the data points, usually a triangulation, and a series of local functions for each element. Bivariate polynomials involve the solution of a single basis, or set of equations, to characterize the entire surface. To summarize, a local transformation is dependent on a subset of nearby data close to the point of evaluation whereas a global method is dependent on all of the data. If the removal of a single datum affects the entire surface, then the method must be considered global in nature.

The nomenclature can be confusing with respect to interpolation and approximation across disciplines. Much of the scattered data literature originates in mathematics and computer science. In a cartographical context, Lam (1983) classifies point (vis-à-vis areal) interpolation as either exact or approximate. This classification carries over into the GIS literature where Burrough (1986) uses the expression "exact interpolator" in the context of interpolation as defined above. The NCGIA Core Curriculum in GIS also uses this terminology (Goodchild and Kemp 1990). If these cases, if a method is not an exact interpolator, then it is an approximating technique.

Back to top

Image Registration with Scattered Data Methods

At a sufficiently large scale, as in the case of typical airborne scanned imagery, the Earth's curvature may be ignored greatly simplifying the distortion modeling. In real Euclidean space, using rectangular Cartesian coordinates, the image registration problem may be separated into two components. Each component consists of fitting a surface to a plane based on a set of control points. The general surface fitting problem may be represented in explicit form as:

$$z = f(u,v) \qquad (1)$$

In this case, the surface value z is expressed as a function of its location in the plane (u,v). The general image warping or geometric transformation problem may be represented in an equation of the form:

$$(x,y) = f(u,v) \qquad (2)$$

This states that the (x,y) input coordinates depend on some function of the (u,v) output coordinates. This is expressed as in inverse transformation as it greatly simplifies the resampling process. A forward transformation is more difficult to implement due to problems in the reconstruction and sampling of the DNs (Wolberg 1990). In inverse warping, the input locations are determined based on functions of the output coordinates and the resampling takes place in the input image. It should be clear that resampling, image reconstruction and sampling does not constitute a form of geometric correction. For example, the following description of "geometric correction" yields little practical information, "the September data were geometrically rectified to the June data using the nearest neighbor algorithm" (Curran and Pedley 1990). The complete image transformation or warping consists of both a geometric correction and a resampling.

The geometric modeling of the input pixel locations may be calculated separately as follows:

$$x = f(u,v) \text{ and } y = f(u,v) \qquad (3)$$

In this form, a wide range of surface fitting methods may be used to model the geometric distortion. Three approaches will be described, each having previously been applied to image warping in recent published literature: (1) inverse distance weighted methods; (2) finite element methods; and (3) bivariate polynomials. Radial basis functions will be discussed in the next section.

In its simplest form, inverse distance weighting is often referred to as Shepard's method (see Franke and Nielson 1991). Its deficiencies are well-known and many modifications have been suggested and reviewed (see Alfeld 1989; Watson 1992). It is a global function that is almost always localized based on the assumption that far away data have little effect on the interpolation. Ironically, the basic methodology survives within the RS/GIS community with a persistence akin to the polynomial approach to image warping. For example, Hodgson, Cheng et al. (1995) apply heuristics and supercomputing power to hurry the construction of this generally unacceptable interpolant. In modified forms, Shepard's method may be suitable for surface fitting and image warping due to its use of ad hoc parameterization (Ruprecht and Muller

1995).

Finite element methods were developed in the engineering literature and have been used for image warping for almost two decades. The simplest form, piecewise linear patches consisting of affine transformations, are constructed based on a triangulation of the control points in the output image. There exist many different triangulation criteria. However, for image warping, no data-dependent or surface-dependent triangulations were located in an extensive literature review.

Two triangulation schemes representing different criteria are the "greedy" (Manacher and Zobrist 1979) and Delaunay (Aurenhammer 1985) triangulations are in The RAND Corporation's CAGIS (Cartographic Analysis and Geographic Information System; Zobrist, Marcellino and Daniels 1991). The Delaunay triangulation has many desirable properties for surface modeling including the avoidance of long narrow triangles and a minimal roughness property in the piecewise linear case (Rippa 1990). In particular, the addition or deletion of a point has only a local effect preventing a rippling throughout the dependent surface as with any other triangulation criteria. The constrained Delaunay triangulation (not strictly a Delaunay triangulation) allows an edge or edges to be fixed (Chew 1989). This makes the triangulation more suitable for applications with "discontinuous" surfaces, e.g. cliffs or cultural features. Adding interior points to a Delaunay triangulation without information on the surface or on nearby points only distorts the rectification of airborne data in an arbitrary manner contrary to research by Devereux, Fuller et al. (1990).

Piecewise linear methods generate surfaces that are continuous but are not differentiable across boundaries, i.e. C0 surfaces. This type of warp has been referred to as tin-sheeting as opposed to rubber-sheeting. C1 surfaces with at least continuous first derivatives across the surface can be more desirable due to a smoother appearance. In remote sensing applications, it will not always be clear if a smoother surface is more accurate as well as being more visually pleasing.

Cubic (Goshtasby 1987) and quintic (Parr and Comer 1990) surface patches are two approaches to improving interpolation quality by including derivative information based on nearby points. The quality of the derivative information greatly influences the surface fit. Data-dependent interpolation is discussed by Dyn and Rippa (1993) and Schumaker (1993) and is becoming increasingly visible.

The methods above interpolate the data. The final method discussed in this section are bivariate polynomials. The following expression is equivalent to an affine transformation when N=1, i.e. a first-order polynomial. The general bivariate formulation may be given as:

$$x = \sum_{i=0}^{N}\sum_{j=0}^{N-i} a_{ij} u^i v^j \text{ and } y = \sum_{i=0}^{N}\sum_{j=0}^{N-i} a_{ij} u^i v^j \qquad (4)$$

where N is the degree or order of the polynomial. These polynomials are equivalent to the trend surface form, although no assumptions are made regarding the existence of any underlying stochastic processes. These polynomials are ill-conditioned, especially at higher orders in which case the solutions must be viewed with skepticism (Ripley 1981). The problem is exacerbated in image-to-map rectification when the output units are quite large, such as might be found in a UTM projection or State Plane Coordinate System.

Modern algorithms are available for solving bivariate polynomial equations such as the Modified Gram-Schmidt, and QR and SVD decompositions (Golub and Van Loan 1989). The pseudoinverse and normal equation approaches should be strictly avoided. Numerical methods are important, especially if RS/GIS investigations are to be conducted under the rubric of science. However, least-squares solutions for bivariate polynomials are well-known and should not require repetition as in Mather (1995). The discussion in Mather on bivariate polynomials perpetuates the almost single-minded focus on bivariate mapping polynomials in remote sensing.

An important and often over-looked characteristic of the least-squares bivariate modeling is its affine invariant form. The coordinate system is arbitrary; the same surface results regardless of scale changes, rotation and translation of the output coordinates. However, the coefficients may be different (Ripley 1981). This information may be used to scale map coordinates, e.g. UTM, and improve the results of the numerical computations. Tensor-product forms as given in Hall (1979) do not possess this property. Similarly deficient is the stepwise selection of parameters for inclusion in the basis (system of equations) if the full trend surface form is not specified.

If the coordinate system arbitrarily affects the surface fit, then any arbitrary origin (translation) and rotation of the output space would effectively change the model. Following Tobler (1994) and Ripley (1981) it would seem prudent to fit coordinate invariant models to the problem or process, deterministic or stochastic. Shepard's method and the greedy and Delaunay triangulations are translation, rotation and scalar invariant, but depending on the triangulation criteria, data-dependent triangulations may be only translation and scalar invariant. In fact, without attention to the physical units in the warp specification, it is possible to generate a myriad of different solutions to the geometric correction problem. This includes local methods such as triangulations as well as global methods. Principally, this refers to expressing coordinates in image space versus map space.

Based on the previous discussion and the survey papers by Franke (1982), Alfeld (1989) and book by Lancaster and Salkauskas (1986), it is possible to identify a set of characteristics that are desirable for scattered data interpolation methods and image warping of remotely sensed data. In particular, the methods must be general--no specific structure to exploit or peculiar configurations to confound--and the data assumed to be relatively accurate. The complexity of the method and its computational requirements must be reasonable. The surface should be smooth--C0 at a minimum and preferably C1 (continuous first derivatives and visually smooth). The method must be coordinate invariant in the sense that the origin shall not influence the surface construction and the implementation of any given scheme must be reproducible. Ideally, the surface should depend continuously on the data. The existence and uniqueness of the solution is presumed. Finally, the effects of free parameters should be understood well-enough to provide acceptable defaults. These criteria are for image warping. The merging of coincident vector coverages, or vector data from a distorted image are different, though related problems. In particular, a forward transformation may be appropriate.

Back to top

Radial Basis Function Methods

Hardy's multiquadrics, and thin plate splines, first appeared in publications on image warping by Gopfert (1982) and Goshtasby (1988), respectively. Both Hardy's multiquadric (MQ) and the thin plate spline (TPS) are global methods and both have been the subject of considerable theoretical and applied research. Other radial basis function methods have been applied to image warping as well, but are not investigate here (see Arad et al. 1994; and Arad and Reisfeld 1995).

In most theoretical work, the functional formulation is very general. For clarity, this presentation is limited to the two-dimensional case. As in the previous section, the surface fitting process is solved separately for image warping in x and y. The radial basis functions may be constructed as a linear combination of the following equations for x and y (presented once for z, an arbitrary variable):

$$\sum_{j=1}^{N} a_j h_j(u_i, v_i) + \sum_{j=1}^{M} b_j q_j(u_i, v_i) = z_i, \quad i = 1, \ldots, N \qquad (5)$$

subject to the following constraints:

$$\sum_{i=1}^{N} a_i q_j(u_i, v_i) = 0, \quad j = 1, \ldots, M \qquad (6)$$

where N is number of GCPs and M is the number of terms in a bivariate polynomial as formulated above (referred to as the polynomial precision of the method). The basis for the MQ and reciprocal multiquadrics (RMQ), may be expressed as follows:

$$h(u,v) = \left(d^2 + c^2\right)^{\pm 1/2} \text{ or } h(r) = \sqrt{\left(d^2 + c^2\right)} \text{ and } h(r) = 1 \Big/ \sqrt{\left(d^2 + c^2\right)} \qquad (7)$$

where d is the distance between (x,y) to every other point and c is the multiquadric parameter. The effect from a point ui to all equidistant points is the same, expressed as a translate of the radial function h. Originally, Hardy (1971) included no polynomial precision in the formulation. The existence and uniqueness of a solution to the multiquadric equations with or without polynomial precision was proven by Micchelli (1986; see Franke and Nielson 1991). The basis functions are formulations of the upper sheet of a hyperboloid of revolution.

The thin plate spline (TPS) is so-called due to its approximation the least bent infinite, infinitely thin plate or lamina under point loads z at locations (u,v). Its formulation may be given as:

$$h(u,v) = d^2 \log d \text{ or } h(u,v) = d^2 \log d^2 \qquad (8)$$

where the latter form simplifies computations (the a coefficients in (8) are scaled by one-half). This method was introduced by Harder and Desmarais (1972); background references are given in Franke and Nielson (1991). The TPS has linear polynomial precision based on its theoretical development. Matrix formulations of the MQ and TPS are given in Ehlers and Fogel (1994) and Bookstein (1989), respectively. The separation of the radial and monomial (polynomial) bases is given in Lancaster and Salkauskas (1986).

Polynomial precision refers to a method's ability to reproduce a polynomial of a given order. Important in other fields, such as computer aided graphics design, and approximation theory, polynomial precision has less significance in remote sensing. Conceivably, polynomial precision could be important for mosaicking or in localized, adaptive variations of radial basis functions. Carlson and Foley (1991) note that unless the surface may be closely approximated by a low-order polynomial, the MQ basis should not be augmented except perhaps by a constant. Franke (1987) and Fogel and Tinney (1996) found that even a constant may lessen the quality of the MQ fit based on the RMSE on known functions or at known check points. Though not studied here, the completely regularized spline of Mitsov and Mits (1993) has a constant trend or constant polynomial precision.

It is well-known that the solutions to radial basis functions can be problematic for large numbers of points--large being application specific. The condition numbers for these systems rise with N and inevitably threaten the computations due to rounding error. For image warping with the TPS, Barrodale, Kuwahara et al. (1993) follow Franke's (1982) comments and scale the data to the unit square [0,1]2 greatly reducing the condition number. Franke and Nielson (1991) report the MQ method has a milder condition number. The scaling strategy is known to allow for the solution of problems involving more than 1000 and less than 2000 points. The computational cost is rather small, and unlike properly scaled bivariate polynomials, the coefficients to the scaled problem may be rescaled back to the original problem space.

Hardy's Multiquadrics

In an extensive study by reviewed in Franke (1982), the multiquadric method proved to be one of the best methods for interpolating over a set of different "known" surfaces from (not too) scattered observations. A review of the method with more than 100 references is given by Hardy (1990). The method has only recently been applied to digital image warping although it finds mention in Wolberg (1990). In computer graphics, criteria for good interpolation are different than in remote sensing applications, however the work by Ruprecht and Muller (1993, 1995) is instructive.

Ruprecht and Muller (1993) place a premium on visual and computational aspects in a comparison of a modified version of Shepard's method, the MQ method and the TPS. The radial basis functions were preferred although the parameter adjustment of the modified Shepard's method was more intuitive. In this paper they also use a variable multiquadric parameter. The existence and uniqueness of a solution is no longer guaranteed.

In subsequent work Ruprecht and Muller (1995) implemented a locally-bounded MQ and implemented another free parameter. This made the resulting surface more "tunable" while increasing the computational efficiency and limited the influence of far away points. Notably, the applications are ad hoc, a characteristic of both Shepard's method and its (improved) variants and the MQ method (see Alfeld, 1989). However, the lack of ready alternatives encourages further study. The effects of the multiquadric parameter on interpolation are reviewed in Hardy (1990) and Carlson and Foley (1991).

Gopfert (1982) specified a two-stage approach to image warping which involves estimating the trend using bivariate polynomials and then applying the MQ method to the residuals (residuals are not errors, per se) thus interpolating the GCPs (see Ehlers and Fogel 1994). Fogel and Tinney (1996) found this approach performed better than if low-order polynomial precision had been included in the system of equations, however they provide no results. Fogel and Tinney also noted that the algorithm may prove less useful at extrapolating beyond the convex hull without at least linear (first-order) precision.

Thin Plate Splines

The thin plate spline has received more attention for image warping than the MQ method. In part, this may reflect the well-understood (hence desirable) minimum-curvature properties. It is certainly more intuitive than the MQ method. Image warping using the TPS in computer graphics (Arad et al. 1994; Arad and Reisfeld 1995; and Ruprecht and Muller 1993); and remote sensing (Goshtasby 1988; Barrodale, Kuwahara et al. 1993); Flusser 1993; and Ehlers and Fogel 1994) has generally found favor. Bookstein (1989) provides a decomposition of the TPS mapping from a morphological viewpoint that complements the image warping research.

Flusser (1993) comments on the computational cost of this method and proposes an adaptive approach that depends on the nature of the distortions. However, properties of the thin plate spline have been exploited by Powell (1992a) and implemented by Barrodale, Skea et al. (1993) dramatically reducing the computational requirements with little loss in accuracy. Depending on the nature of the distortions, it appears that the method requires many control points to be effective.

Barrodale's (1995) implementation of Powell's (1992a) scheme transforms entire skeletons or arbitrary shapes as well. This effectively extends the control point-based TPS warp to what may loosely referred to as a feature-based method. However, it differs greatly from mesh-based warping (see Wolberg, 1990) and feature-based schemes (Beier and Neely 1992) in computer graphics. The utility for RS/GIS data integration is obvious given the difficulties in locating coincident points distributed in such a manner as to accurately characterize the distortions.

Bookstein and Green (1993) and Bookstein (1995) build on properties of the thin plate spline to characterize biological shape (biometrics; morphometrics) to include derivative information, called edgels, that may be further used to further specify the deformations. Arad (1994) and Arad and Reisfeld (1995) provide some indication that grid-based patches may be used to refine the warp by coupling the deformation calculations of (u,v) and (x,y). It is not clear to what extent these methods will affect fast evaluation of radial basis function methods such as given in Powell (1992a).

Back to top

Rocky Flats Environmental Technology Site

The motivation for alternative scattered data warping methods follows from the operational shortcomings of the two most commonly used warping methods: bivariate mapping polynomials and piecewise linear finite elements. For many applications, these methods are adequate. However, in a general sense, high spatial resolution airborne scanner remotely sensed data over rapidly varying terrain exposes these models to criticism. The environmental restoration of Rocky Flats provides some context for discussion. The scene and data are described in general terms since the principal goal in this section is to communicate the nature of the rectification problem.

The Rocky Flats facility in Colorado was formerly engaged in nuclear components production. The site is characterized by various contamination problems including, but not limited too ground water contamination. The integration of high spatial resolution remotely sensed data into a GIS for environmental monitoring has a sense of urgency about it. Site specific information and the location of Rocky Flats is shown in Figure 1. More information may be found at the web site for the U.S. Department of Energy Office of Environmental Management (1995).

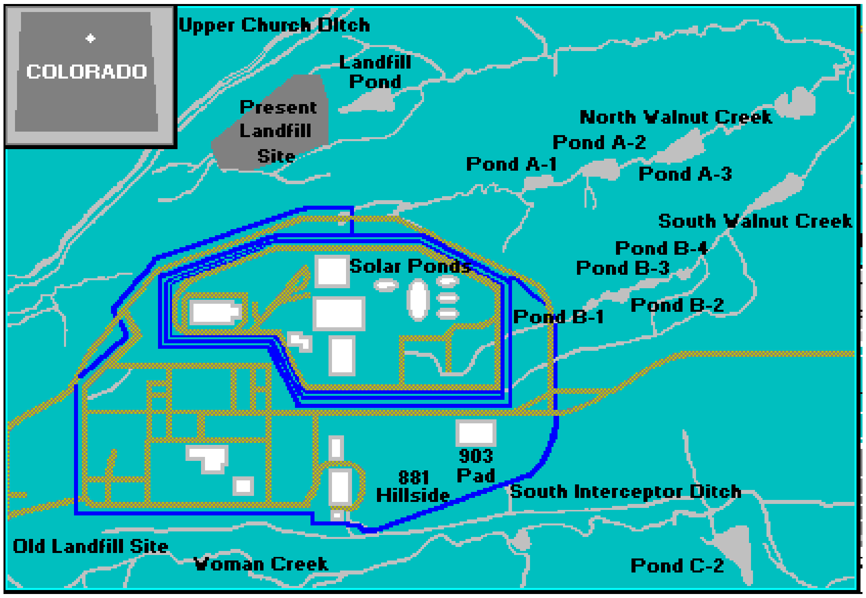


Figure 3: Rocky Flats is located approximately 16 miles northwest of Denver, Colorado (North is toward the top of the page).

The site occupies eleven square miles including a buffer surrounding the operating facilities. The elevation varies from over 6200 feet in the southwest to approximately 5600 feet in the southeast corner. Outside the plant inner boundaries, it is difficult to find adequate ground control for registration or rectification. Rectifying data sampled to less than 5 meter-squared to the local Colorado State Plane Coordinate System using bivariate mapping polynomials has a lack of fit at both control points and check points. The ground control is this case is a 2-feet-squared digital orthophoto of the facility and surrounding area. Piecewise finite element methods improve upon the polynomials provided that the GCPs are relatively accurate and representative of scene variations and the tessellation is similarly representative.

In an application using 98 control points and 24 check points, the radial basis function methods provided a better fit compared to polynomials based on two criteria: RMSE and visual inspection. Finite element methods were not included in this comparison due to time and space constraints. Furthermore, it is thought that implementing data dependent triangulations will provide better results and research should be focused in this direction. The polynomial RMSE results for control and check points are presented in Tables 1 and 2, respectively. The bivariate polynomials were applied only to the third order for two reasons: (1) polynomials are prone to oscillate widely between observations for higher degrees, and (2) polynomials are a distraction from research into better alternative techniques.

Polynomial Degree	RMSE *x*	RMSE *y*	RMSE *total*
First	2.785	7.710	8.197
Second	2.493	2.502	3.532
Third	2.064	1.856	2.776

Table 1 : Bivariate mapping polynomials. Control Points

Table 1: Bivariate mapping polynomials. Control Points

Polynomial Degree	RMSE *x*	RMSE *y*	RMSE *total*
First	3.169	7.061	7.740
Second	3.077	2.848	4.193
Third	2.552	2.008	3.247

Table 2: Bivariate mapping polynomials. Check Points.

Table 2: Bivariate mapping polynomials. Check Points

The implementation of Hardy's multiquadric (Ehlers and Fogel 1994) is a two-stage procedure usually involving the removal of a low order trend and the specification of the multiquadric parameter. Currently, an approximation of the best multiquadric parameter is based on an approximation by Gopfert (1982). The results for the two radial basis functions described previously are presented in Table 3. There are no residuals for these two methods since they are interpolating functions.

Method	RMSE *x*	RMSE *y*	RMSE *total*
MQ	1.303	1.272	1.821
TPS	1.351	1.283	1.863

Table 3: Radial basis function methods. Hardy's multiquadric (first order polynomial, Göpfert's *G*=1.10) and the thin plate spline.

Table 3: Radial basis function methods. Hardy's multiquadric (first order polynomial, Gopfert's G=1.10) and the thin plate spline.

The grouping of observations into control and check points is useful for estimating the relative accuracy of the geometric transformations. This measure of accuracy is relative to how well the distortions have been sampled while it also reflects the adequacy of the model for these points. A useful exercise in evaluating both the adequacy of the sampled distortions as well as the overall model fit is ordinary cross validation. A conscience effort is made to avoid the use of the terms "bias" and "predictors" so that this is not confused with statistical methodologies (see Goodall, p. 494). In this case, the "projection" matrices for the polynomials (Hadi 1996) and the matrices for radial basis functions (see Golub, Heath and Wahba 1979) may be readily used to invoke "leave one out" cross validation. In the former case, the calculations are trivial. For radial basis functions, the computations are more intensive and problematic for large numbers of control points. Generalized cross validation is an alternative yet to be explored in this image warping context (for information on generalized cross-validation, see Craven and Wahba 1979 and Bates, Lindstrom, Wahba and Yandell 1987).

The results of ordinary cross validation based a single calculation of the "projection" or "influence" matrices are shown in Table 4. This is the same result one would get by deleting each observation in turn and estimating the model fit without that observation. This cross validated RMSE still suffers as a metric given that, among other things, some areas may have little ground control and consequently reflect poorly on overall model fitness. In general, it is recommended that the ordinary cross validation method be used primarily to identify problems with geometric models and not solely to establish accuracy summaries that are all too often misleading.

Method	RMSE		
	x	*y*	*total*
Polynomial	2.364	2.103	3.164
Hardy's MQ	1.535	2.270	2.741
Thin Plate Spline	1.482	1.739	2.285

Table 4: Ordinary Cross Validation. Third degree polynomial, Hardy's multi-quadric (first order polynomial, Göpfert's G=1.10) and the thin plate spline.

Table 4: Ordinary Cross Validation. Third degree polynomial, Hardy's multiquadric (first order polynomial, Gopfert's G=1.10) and the thin plate spline.

An example of the multiquadric method applied to this data set using all points for control is shown in Figure 2. The corrected image is shown with some ancillary data included for visualization (i.e. no comments are made here with respect to the data quality). The image was acquired using a Daedalus 1268 MSS scanner. It is an RGB composite of the thermal band (inverted), an infrared band and a portion of the red part of the electromagnetic spectrum. PC-based ArcView 2.1 (ESRI 1995) is used to display the image and other features.

Figure 4: Rocky Flats Environmental Technology Site. 25 foot contours and roads from GIS coverages are presented over airborne scanned image data (see text).

It is reasonable to question the effects of improved registration on subsequent analyses though this is a much tougher task than the analysis of misregistration effects such as in the work cited previously. RS/GIS integrated analyses will improve measurably in amount and quality. Figures 5 and 6 show two rectified areas from the region using PCIs Image Handler to illustrate the potential of new methodologies.

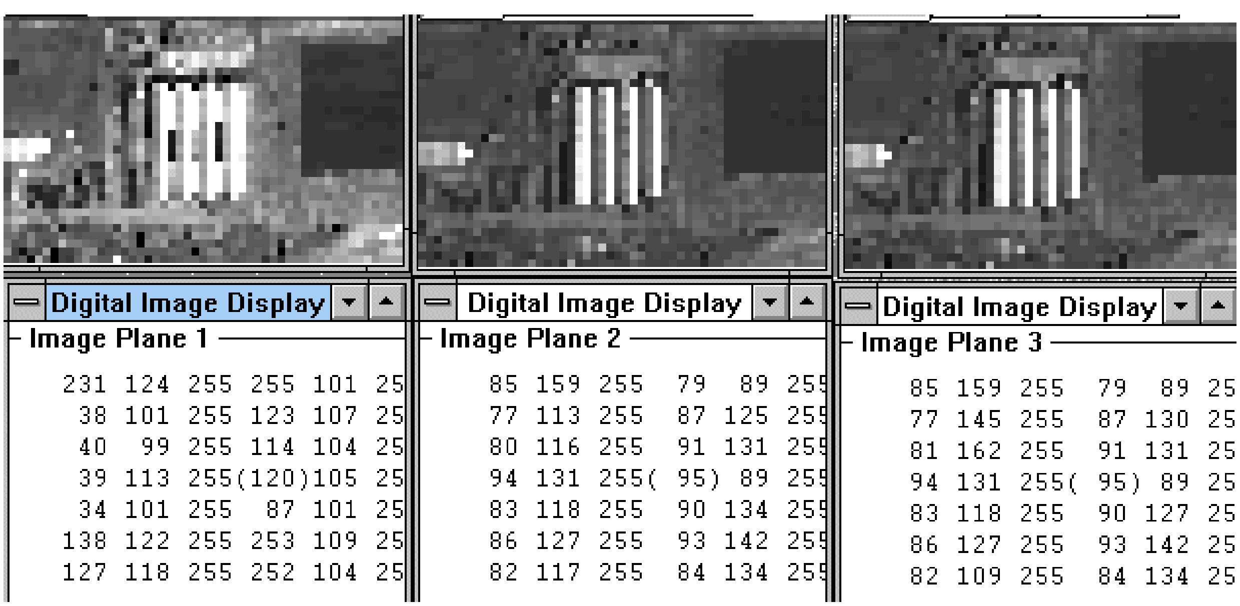


Figure 5: Left-to-right. Third degree polynomial, Hardy's multiquadric (first order polynomial,

Gopfert's G=1.10) and the thin plate spline using nearest neighbor resampling.

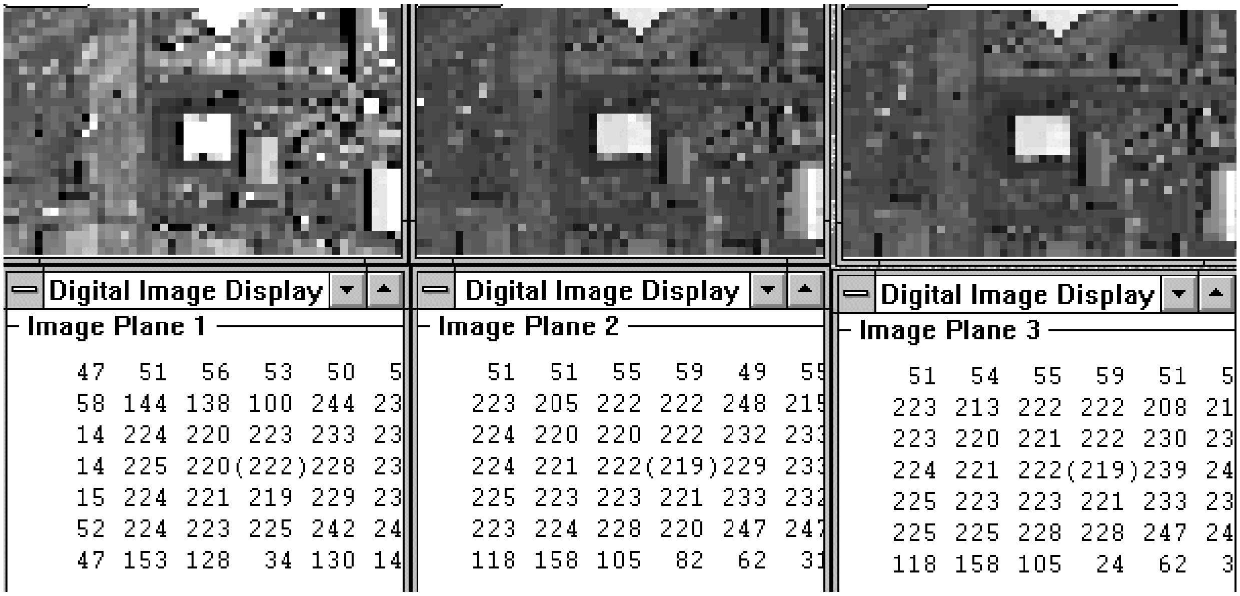


Figure 6: Left-to-right. Third degree polynomial, Hardy's multiquadric (first order polynomial, Gopfert's G=1.10) and the thin plate spline using nearest neighbor resampling.

Back to top

A Software Implementation

The radial basis function methods described here were implemented in several stages. The current software is based on source code for a program called MQE and written by Flottemesch (1993). The program was continually refined over the next sixteen months. Stable and reliable numerical procedures were implemented and some conceptual shortcomings were corrected. At this point, the software was more widely tested. In its current manifestation, MQReg (Multiquadric Registration) uses Flottemesch's code for the user-interface and file-handling. Otherwise it is quite different hence the new name.

The software and a manual will soon be made available via anonymous ftp from the NCGIA (Fogel 1996). Additionally, a revision of a report written is available as an NCGIA Technical Report 96-01 (Fogel and Tinney 1996). The technical report and Ehlers and Fogel (1994) provide a more thorough presentation of radial basis functions and include many more citations.

The program is research code. It is operational, but hardly equivalent to a commercial application. Most notably, it has been structured to be maintainable by others than the programmers. Of course, this means that it must have been written in the C programming language.

At this time, PCI has implemented the thin plate spline (v. 6.0) and ERDAS has been also working on implementing these methods in their commercial packages. Barrodale (1995) has implemented and developed a working demonstration of Powell's (1992a) fast evaluation of thin plate splines. The advanced image rectification software, Spider Warp, can match and warp vector-based objects for highly accurate RS/GIS data integration (Barrodale 1995).

It is expected that more computational details and applications information will be presented in the very near future (Ehlers and Fogel 1996, and McGwire and Fogel 1996). In any case, additional technical information will be deferred until after the general release of this program in the Spring of 1996.

Back to top

Conclusions

The geometric correction of remotely sensed imagery in the absence of ephemeris platform data has been largely ignored by researchers and application specialists. During this period, significant progress has been made in approximation theory, scattered data modeling and surface fitting. The potential utility of these methods is rather obvious, especially given increasingly high spatial resolution data being acquired now and in the future.

Purposely, this presentation has been quite critical of two areas of research: (1) the unfortunate and unwise practice of all but dismissing geometric distortions and their implications in research, and (2) the almost singular focus on bivariate mapping polynomials. This may be regarded as a disservice to those outside the academic and research communities.

The radial basis function methods are intriguing and clearly worthy of increased attention. As Powell (1992b) states, "...the author believes radial basis function methods are becoming as important as piecewise polynomials although good software is not yet available for general computer calculations." It is hoped that this paper will stimulate interest and perhaps provoke debate on this subject within the remote sensing and GIS communities.

Back to top

Acknowledgments

This research was supported in part by an EG&G Energy Measurements Contract (P.O. 27381) and a NASA Office of University Affairs Grant NAGW-1743 (J.E. Estes, Principal Investigator). Larry Tinney, Michael Howard and Lynne Christel of Bechtel Nevada provided valuable input and data for this paper and Ken McGwire (Desert Research Institute, Reno, NV), and Phil Townsend and Prof. Aaron Moody (University of North Carolina exercised the software used in this paper for their own applications helping to validate these results and making the program more robust. Kim Rauenzahn of TASC deserves all the credit for helping arrange this paper in HTML format (and none of the blame for its current state).

Back to top

References

Alfeld, P. (1989) Scattered data interpolation in three or more variables. *Mathematical methods in computer aided geometric design*, T. Lyche and L. L. Schumaker, eds., San Diego: Academic

Press, pp. 1-34.
Arad, N., Dyn, N., Reisfeld, D., and Yeshurin, Y. (1994) Image warping by radial basis functions: application to facial expressions. *CVGIP: Graphical Models and Image Processing* 56(2): 161-172.
Arad, N. (1994) Designing and implementing a grid-distortion mapping based on variational principles. *Computer Graphics Forum* 13(3): C-259-C-270.
Arad, N., and Reisfeld, D. (1995) Image warping using few anchor points and radial functions. *Computer Graphics Forum* 14(1): 35-46.
Aurenhammer, F. (1991) Voronoi diagrams--A survey of a fundamental geometric structure. *ACM Computing Surveys* 23(3): 344-405.
Barrodale, I., Kuwahara, R., Poeckert, R., and Skea, D. (1993) Side-scan sonar image processing using thin plate splines and control point matching. *Numerical Algorithms* 5: 85-98.
Barrodale, I., Skea, D., Berkley, M., Kuwahara, R., and Poeckert, R. (1993) Warping digital images using thin plate splines. *Pattern Recognition* 26(2): 375-376.
Barrodale, I. (1995). SpiderWarp. http://www.barrodale.com/~Ubarro. Victoria, British Columbia: Barrodale Computing Services.
Bates, D.M., Lindstrom, M.J., Wahba, G., and Yandell, B. (1987) GCVPACK - Routines for generalized cross-validation. *Communications in Statistics. B. Simulation and Computation* 16(1): 263-297.
Beier, T., and Neely, S. (1992) Feature-based image metamorphosis. *Computer Graphics* 26(2): 35-42.
Bernstein, R. (1983) Image geometry and rectification. *Manual of Remote Sensing, Second Edition*, R. N. Colwell, ed., Falls Church, Virginia: American Society of Photogrammetry, pp. (Chapter 21) 873-922.
Billingsley, F.C. (1982) Modeling misregistration and related effects on multispectral classification. *Photogrammetric Engineering and Remote Sensing* 48(3): 421-430.
Bookstein, F.L. (1989) Principal warps: thin plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(6): 567-585.

Bookstein, F.L., and Green, W.D.K. (1993) A feature space for edgels in images with landmarks. *Journal of Mathematical Imaging and Vision* 3: 231-261.
Bookstein, F.L., and Green, W.D. (1993) A thin-plate spline for deformations with specified derivatives. *Mathematical Methods in Medical Imaging II*, San Diego, CA, USA, 07/11 - 07/16/93, SPIE, vol. 2035, pp. 14-28.
Bookstein, F.L. (1995) How to produce a landmark point: the statistical geometry of incompletely registered images. *Vision Geometry IV*, San Diego, CA, USA, 07/09 - 07/14/95, SPIE -- The International Society for Optical Engineering, vol. 2573, pp. 266-277.
Brown, L.G. (1992) A survey of image registration techniques. *ACM Computing Surveys* 24(4): 325-376.
Burrough, P.A. (1986) *Principles of geographical information systems for land resources assessment*. Oxford: Clarendon Press.
Carlson, R.E., and Foley, T.A. (1991) The parameter R2 in multiquadric interpolation. *Computers & Mathematics with Applications* 21(9): 29-42.
Chew, L.P. (1989) Constrained Delaunay triangulations. *Algorithmica* 4: 97-108.
Christensen, E.J., Jensen, J.R., Ramsey, E.W., and Mackey, H.E., Jr. (1988) Aircraft MSS data registration and vegetation classification for wetland change detection. *International Journal of Remote Sensing* 9(1): 23-38.
Craven, P., and Wahba, G. (1979) Smoothing noisy data with spline functions. *Numerische Mathematik* 31: 377-403.

Curran, P.J., and Pedley, M.I. (1990) Airborne MSS for Land Cover Classification II. *Geocarto International* 2: 15-26.

Devereux, B.J., Fuller, R.M., Carter, L., and Parsell, R.J. (1990) Geometric correction of airborne scanner imagery by matching Delaunay triangles. *International Journal of Remote Sensing* 11(12): 2237-2251.

Duggin, M.J., and Robinove, C.J. (1990) Assumptions implicit in remote sensing data acquisition and analysis. *International Journal of Remote Sensing* 11(10): 1669-1694.

Dyn, N., and Rippa, S. (1993) Data-dependent triangulations for scattered data interpolation and finite element approximation. *Applied Numerical Mathematics* 12(1-3): 89-105.

Ehlers, M., and Fogel, D.N. (1994) High-precision geometric correction of airborne remote sensing revisited: the multiquadric method. *Image and Signal Processing for Remote Sensing*, Rome, Italy, 26-30 September 1994, SPIE, vol. 2314, pp. 814-824.

Ehlers, M., and Fogel, D.N. (1996) Radial basis function methods for image rectification: background and implementation (in preparation). ERDAS. (1995) *ERDAS/Imagine 8.1.* , ERDAS, Inc., Atlanta, Georgia.

ESRI. (1995) *ESRI Arcview 2.1.* , ESRI, Redlands, California.

Estes, J.E., and Star, J.L. (1993) *Remote Sensing and GIS Integration: Towards a Prioritized Research Agenda.*, Technical Report Number 93-04, National Center for Geographic Information and Analysis, University of California, Santa Barbara.

Fogel, D.N., and Tinney, L. (1996) *Image registration using multiquadric functions.*, Technical Report Number 96-01, National Center for Geographic Information and Analysis, University of California, Santa Barbara.

Forshaw, M.R.B., Haskell, A., Miller, P.F., Stanley, D.J., and Townshend, J.R.G. (1983) Spatial resolution of remotely sensed imagery: a review paper. *International Journal of Remote Sensing* 4: 497-520.

Franke, R., and Nielson, G. (1980) Smooth interpolation of large sets of scattered data. *International Journal for Numerical Methods in Engineering* 15: 1691-1704.

Franke, R. (1982) Scattered data interpolation: tests of some methods. *Mathematics of Computation* 38(157): 181-200.

Franke, R. (1987) Recent advances in the approximation of surfaces from scattered data. *Topics in Multivariate Approximation*, C. K. Chui, L. L. Schumaker, and F. I. Utreras, eds., Orlando, Florida: Academic Press, Inc., pp. 79-98.

Franke, R., and Nielson, G. (1991) Scattered data interpolation: a tutorial and survey. *Geometric modelling: methods and applications*, H. Hagen and D. Roller, eds., Berlin: Springer-Verlag, pp. 131-160.

Golub, G.H., and Loan, C.F.V. (1989) *Matrix computations, second edition*. Baltimore, Md.: Johns Hopkins University Press.

Goodall, C.R. (1993) Computation using the QR decomposition. *Computational statistics*, C. R. Rao, ed., New York: North-Holland, pp. 467-508.

Goodchild, M.F., and Kemp, K.K., eds. (1990) *NCGIA Core Curriculum: Technical Issues in GIS*. Santa Barbara, California, 93106: National Center for Geographic Information and Analysis University of California, Santa Barbara.

Goshtasby, A. (1987) Piecewise cubic mapping functions for image registration. *Pattern Recognition* 20(5): 525-533.

Goshtasby, A. (1988) Registration of images with geometric distortion. *IEEE Transactions on Geoscience and Remote Sensing* 26(1): 60-64.

Gopfert, W. (1982) Methodology for thematic image processing using thematic and topographic data bases and base-integrated multi-sensor imagery. *Proceedings, ISPRS Commission VII Symposium*, Toulouse, France, vol. I, pp. 13-19.

Hadi, A.S. (1996) *Matrix algebra as a tool*. Belmont, California, Duxbury Press.
Hall, E.L. (1979) *Computer image processing and recognition*. New York: Academic Press.
Harder, R.L., and Desmarais, R.N. (1972) Interpolation using surface splines. *Journal of Aircraft* 9: 189-191.
Hardy, R.L. (1971) Multiquadric equations of topography and other irregular surfaces. *Journal of Geophysical Research* 76(8): 1905-1915.
Hardy, R.L. (1990) Theory and applications of the multiquadric-biharmonic method . *Computers and Mathematical Applications* 19 (8/9): 163-208 .
Heard, M.I., Mather, P.M., and Higgins, C. (1992) GERES: a prototype expert system for the geometric correction of remotely-sensed images. *International Journal of Remote Sensing* 13(17): 3381-3385.
Hodgson, M.E., Cheng, Y., Coleman, P., and Durfee, R. (1995) Computational GIS burdens. *Geo Info Systems* 5(4): 28-37.
Lam, N.S.-N. (1983) Spatial interpolation methods: a review. *The American Cartographer* 10(2): 129-149.
Lancaster, P., and Salkauskas, K. (1986) *Curve and surface fitting: an introduction*. New York: Academic Press.
Manacher, G., and Zobrist, A.L. (1979) Neither the greedy nor the Delaunay triangulation of a planar point set approximates the optimal triangulation. *Information Processing Letters* 9(1): 31-34.
Marble, D.F., and Peuquet, D.J. (1983) Geographic information systems and remote sensing. *Manual of Remote Sensing, Second Edition*, R. N. Colwell, ed., Falls Church, Virginia: American Society of Photogrammetry, pp. (Chapter 22) 923-958.
Markarian, H., Bernstein, R., Ferneyhough, D.G., Gregg, L.E., and Sharp, F.S. (1973) Digital correction for high resolution images. *Photogrammetric Engineering* 39: 1311-1320.
Mather, P.M. (1995) Map-image registration accuracy using least-squares polynomials. *International Journal of Geographical Information Systems* 9(5): 543-554.
McGwire, K. and Fogel, D.N. (1996) Radial basis function methods for image rectification: methdology and applications (in preparation). Micchelli, C.A. (1986) Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation* 2(1): 11-22.
Mitsov, H., and Mits, L. (1993) Interpolation by regularized spline with tension: I. Theory and implementation . *Mathematical Geology* 25 (6): 641-655 .
NCGIA. (1991) *Initiative 12: Integration of Remote Sensing and Geographic Information Systems, Report of the Specialist Meeting*, Technical Report Number 91-16, National Center for Geographic Information and Analysis, University of California, Santa Barbara.
Niblack, W. (1986) *An introduction to digital image processing*. Englewood Cliffs, N.J.: Prentice-Hall International.
Parr, J.T., and Comer, R.P. (1990) A local area geometric warping algorithm. *ACSM/ASPRS Annual Convention*, Denver, Colorado, 18-23 March 1990, ASPRS, vol. 4, pp. 316-320.
Powell, M.J.D. (1992a) Tabulation of thin plate splines on a very fine two-dimensional grid. *Numerical methods of approximation theory*, D. Braess and L. L. Schumaker, eds., Basel: Birkhuser-Verlag, pp. 105-210.
Powell, M.J.D. (1992b) The theory of radial basis function approximation in 1990. *Advances in Numerical Analysis*, W. Light, ed., Oxford: Oxford Science Publications, pp. 105-210.
Ripley, B.D. (1981) *Spatial Statistics*. New York: John Wiley.
Ruprecht, D., and Muller, H. (1993) Free form deformations with scattered data interpolation methods. *Geometric Modelling (Computing Supplement 8)*, G. Farin, H. Hagen, and H. Noltemeier, eds., Wien, Germany: Springer-Verlag, pp. 267-281.

Ruprecht, D., and Muller, H. (1995) Image warping with scattered data interpolation. *IEEE Computer Graphics and Applications* 15(2): 37-43.

Schumaker, L.L. (1993) Computing optimal triangulations using simulated annealing. *Computer Aided Geometric Design* 10: 329-345.

Singh, A. (1989) Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing* 10(6): 989-1003.

Star, J.L., ed. (1991) *The integration of remote sensing and geographic information systems*. Bethesda, Maryland: American Society of Photogrammetry and Remote Sensing.

Thormodsgard, J.M., and Lillesand, T. (1987) Comparison of the gridded finite element and the polynomial interpolations for geometric rectification and mosaicking of Landsat data. *Proceedings of the ASPRS-ACSM Annual Convention*, Baltimore, Maryland, vol. 2, pp. 139-151.

Townshend, J.R.G., Justice, C.O., Gurney, C., and McManus, J. (1992) The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing* 30(5): 1054-1060.

Trotter, C.M. (1991) Remotely-sensed data as an information source for geographical information systems in natural resource management. *International Journal of Geographical Information Systems* 5(2): 225-239.

U.S. Department of Energy. (1995) http://www.em.doe.gov/. U.S. Department of Energy Office of Environmental Management: U.S. Department of Energy Office of Environmental Management.

Van Huffel, S., and Zha, H. (1993) The total least squares problem. *Computational statistics*, C. R. Rao, ed., New York: North-Holland, pp. 377-408.

Wivell, C. (1995). http://geo.arc.nasa.gov/esd/esdstaff/landsat/landsat.html. Sioux Falls, South Dakota: EROS Data Center.

Wolberg, G. (1990) *Digital Image Warping*. Los Alamitos, California: IEEE Computer Society Press.

Zobrist, A.L., Marcellino, L.J., and Daniels, G.S. (1991) *RAND's Cartographic Analysis and Geographic Information System (RAND-CAGIS): A Guide to System Use*. Santa Monica, CA: The Rand Corporation.

Back to top

Author Information

David N. Fogel
Research Fellow
National Center for Geographic Information and Analysis
University of California, Santa Barbara, CA 93106-4060
E-mail:fogel@ncgia.ucsb.edu

Back to top

Arne Koschel, **Ralf Kramer**, **Ralf Nikolai**, **Wilhelm Hagg**, **Joachim Wiesel**, **Heiko Jacobs**

A Federation Architecture for an Environmental Information System incorporating GIS, the World Wide Web, and CORBA

Abstract:

In environmental information systems two important problems have to be solved. First, environmental questions touch upon a large number of fairly different knowledge domains. These questions are normally addressed by different systems, which are usually heterogeneous, both technical and semantical. Therefore, in order to answer environmental questions, e.g., in state authorities, data and functionality from different sources have to be available uniformly. Second, environmental information kept by authorities should be available for the public as well.

Nowadays, emerging standards for open, distributed systems like the Object Management Group's Common Object Request Broker Architecture cope with the first problem. Especially on a global scale, the World Wide Web (WWW) and its tools tackle the second problem.

For these purposes, a WWW-based application has been developed that enables clients without specific software tools to get geographic information and business graphics for visualizing tabular data. Maps and graphics can be combined to create thematic maps or to down-load maps and graphics for further processing at the client site. In a first implementation, we use the GIS GRASS and GNUPLOT with some extensions as services to provide map and business graphics pictures. In this paper we present the implementation of our GIS access services and we show how this fits nicely into our federation architecture.

1 Introduction

In large environmental information systems (EIS) two important problems have to be solved. First, environmental questions touch upon a large number of fairly different knowledge domains. These questions are normally addressed by different autonomous systems, which are usually quite heterogeneous, both technical (e.g., different operating systems, programming languages etc.) and semantical (e.g., different data models). Therefore, in order to answer environmental questions, e.g., in state authorities, data and functionality from different sources have to be available uniformly. Second, environmental information kept by authorities should be available for the public as well.

Nowadays, emerging standards for open, distributed systems like the Object Management Group's (OMG) Common Object Request Broker Architecture (CORBA) cope with the first problem. Especially on a global scale, the World Wide Web (WWW) and its tools tackle the second problem.

A focal point in EIS are Geographic Information Systems (GIS), which provide geographic data and maps for environmental modelling and statistical data analysis. Integrating GIS objects is a difficult task, since there are several standards for data formats and a wide field of methods for data analysis. A standardization of data formats is currently under way within the OGIS (Open Geodata Interoperation Specification) effort. However, no standard is available yet.

The remainder of this paper is organized as follows: In Section 2 we describe the federation architecture, which is based on the World Wide Web and CORBA. Section 3 describes in some more detail how GIS functionality and visualization facilities are implemented. Section 4 elaborates on using CORBA for accessing relational databases, which are among the most important information sources in EIS. A survey on related work is given in Section 5. Section 6 presents the conclusions and gives an outlook on future work.

2 Outline of the Federation Architecture

Figure 1 depicts a high-level view of the architecture. We assume that the reader is familiar with the World Wide Web (WWW). A short introduction to WWW and the Common Gateway Interface is given in [McCauley 96].

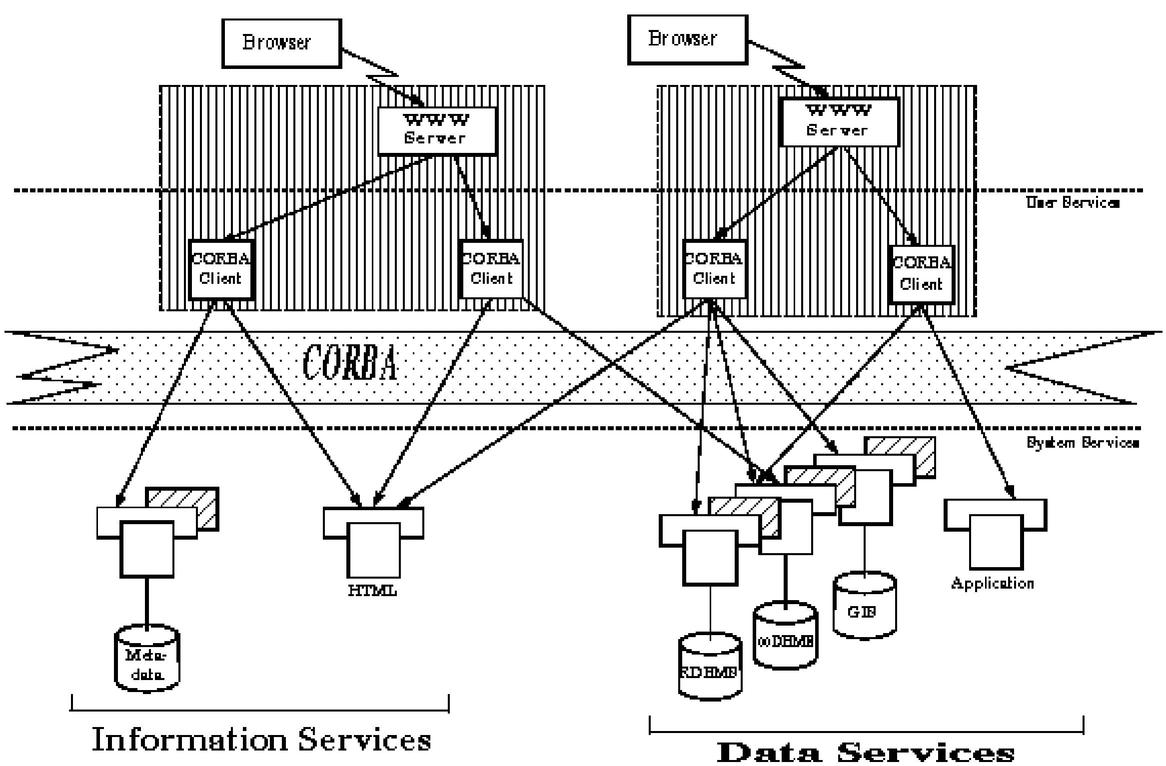


Figure 1: Overall Software Architecture

Horizontally, we distinguish between application level and system level services. System level services provide basic functionality such as accessing databases and preparing HTML pages. Application level services basically combine several system level services into higher level services that are available for the user. Using CORBA, these services can be transparently assigned to computers in a network.

Vertically, we distinguish between information and data services. The information services help the user to find relevant data sources, e.g., data services and reports. They are based on the environmental data catalogue UDK. This metainformation system is used in most German states and in Austria. For an in-depth description of the current development of the UDK we refer to [Lessing 96]. Our first prototype made UDK data available using the tools and techniques of the World Wide Web [Kramer 95,Kramer 96]. Meanwhile, we extended the underlying UDK data model in order to provide online access to data services as shown in figure 1. These data sources primarily are environmental databases and reports that are as well accessible with WWW tools. Data sources may comprise geographic information systems (GIS), relational, pre-relational, object-oriented database systems, and expert systems. Data services deliver results that in turn are used as input for the visualization facilities described in Section 3.

3 Visualization Facilities

3.1 Requirements

Visualization of datasets is an essential task in all information systems. Requirements to visualization tools increase with the complexity of the data. In EIS, it is absolutely mandatory to handle both tabular data and complex geometric data; tabular data have georeferences as well. Both data types have to be visualized such that both an overview of all data and -- on user's request -- details of the data are presented.

Several GIS provide such facilities. An overview of existing GIS can be found in [Rodcay (ed.) 95]. However, these GIS are fairly complex to handle. Hence, they are inappropriate for the casual user. Furthermore, these GIS consume huge amount of resources (e.g., main memory) and are pretty expensive.

3.2 General Approach and Tool Selection

In the architecture of figure 1, we use a GIS and other visualization tools at the WWW-server site to meet those requirements. At the client site, only a standard WWW browser is required. All tools used are in the public domain; hence, there are no licence fees incured at the server site. In the remainder of this Section, we describe those tools and the techniques to connect them to WWW.

The implementation of our services is based on Tcl [Ousterhout 94], a high level command interpreter that provides easy access to files and convenient string manipulation. Tcl as well offers many different data types such as lists and associative arrays.

In order to extract maps from the GIS and to provide complex GIS functionality, we use the public domain tool GRASS, which is a raster oriented GIS that supports digital image processing, map creation and several vector operations. GRASS is implemented in C and is available on several UNIX platforms. Using a public domain GIS enables source code level debugging and the extension of the existing pool of methods, since the source code is available [U.S. Army Corps of Engineers 95]. In addition, the GRASS mailing list is a good forum for the discussion of problems [GRASS 95].

The visualization of tabular data requires a tool for generating business graphics. Currently, we use GNUPLOT [GNU 95] to create line-, bar-, and pie-charts. Since GNUPLOT is originally a program designed for scientists to plot data and mathematical functions, it does not match the requirements exactly. To draw bar charts, e.g., is not very effective. Extensions to GNUPLOT are restricted to a given version of GNUPLOT; there is no guaranty for a support in the future. Other programs, which might match the requirement better, are to our knowledge currently not available in the public domain.

In addition, we need a tool to combine outputs from different raster data sources such as maps, diagrams and icons, and for different format conversions between standard raster image formats like GIF, PPM and TIFF. We use Tom Boutell's gd-library [Boutell 95] for operations on GIF-images. The FLY-tool [Gleeson 95] from Martin Gleeson, which is a commandline interface to the gd-library, provides easy access to library functions within TCL. The pbmplus tools [PBMPLUS 95] are used for format conversions.

3.3 GIS Functionality

A GIS has to provide detailed information of a geographic region. We split the GIS functionality into two parts: visualization of GIS data on the one, data manipulation and analysis on the other hand. Visualization is done by a server GIS, which is connected via CGI [McCool 94] to the WWW-server (see figure 2). Currently, we use GRASS. Further services, provided by individual GIS at the WWW server site, are to be provided by data transfer between these GIS and the server GIS. This enables the integration of several GIS methods, using a simple transfer interface between standard GIS data formats. Up to now, an offline interface to transfer data to the server GIS has been developed and implemented.

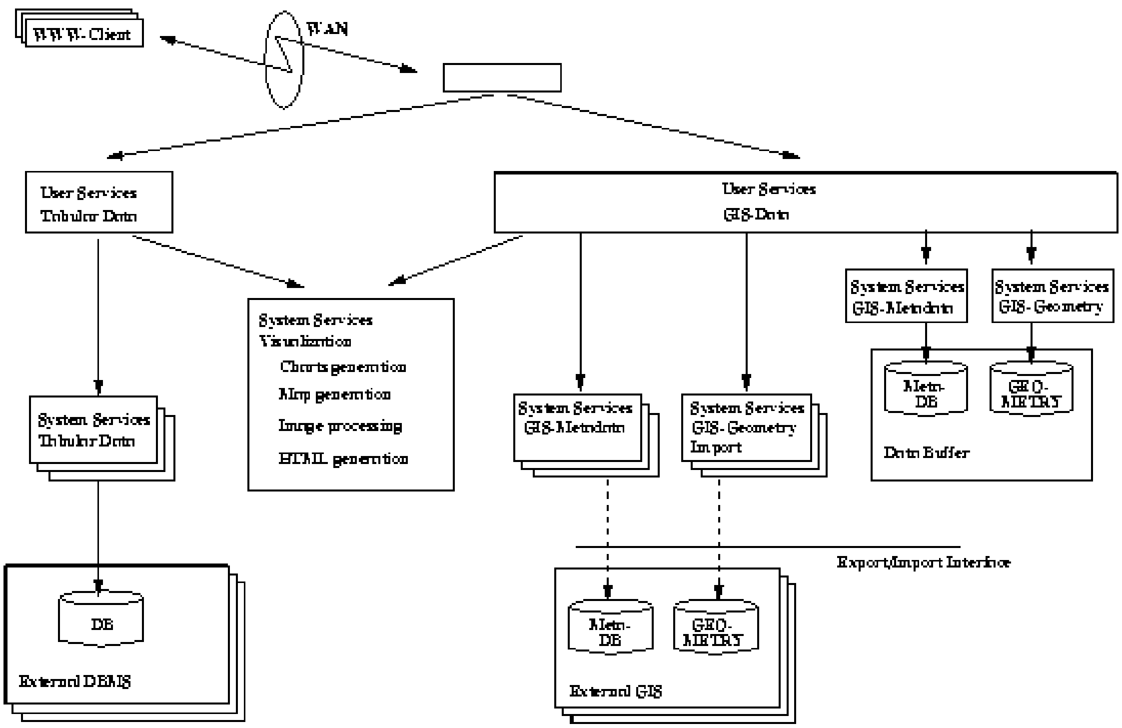


Figure 2: Architecture of GIS Integration

The architectural details for handling GIS data are shown in figure 2. We have WWW clients, the WAN, and the WWW server. Then, we have application level services, i.e., services for tabular data and GIS services. According to the overall architecture as described in Section 2, application level services use different system level services to fulfill their complex tasks. The latter can be divided into three groups:

1. services for different DBMS on different platforms as shown in the left branch of figure 2,
2. services for GIS data as shown in the right part of figure 2, and
3. services for visualization, which include extraction of maps, generation of business graphics, and a combination of both in single HTML result pages.

GIS services are further divided into services for geometric data and metadata. Complex geometric objects are stored in individual GIS formats, whereas metadata of each geometric object are typically stored in a relational DBMS. In our architecture, GIS data are imported into the server GIS. Hence, the server GIS can as well be seen as a buffer that stores data for a limited period of time. Since metadata and geometric data are closely interrelated, for obvious reasons, metadata is stored in this buffer as well. Thus, we have system services that provide data buffered locally to the user services and, on the other hand, services to import and export the data to several GIS. This approach makes both data and methods -- originating in individual GIS -- available on the Web.

Users can choose the region to be shown, different map layers to be displayed in a combination, and the style, in which the map and its elements appear in the final representation. Maps are separated into background and foreground maps. Background maps contain information for all points of the image. Foreground maps are any kind of vector or site maps. It is possible to combine one background with any number of foreground maps. Zoom facilities are provided for the maps. Furthermore, the user can mark a region and use this region for subsequent database queries, e.g., queries for UDK data.

3.4 Business Charts

To get familiar with the content of extended tabular data sets, a representation as business graphic is required. Since nearly all environmental data sets have a georeference, it is neccessary to visualize the georeference as well. Therefore, we have developed a method to combine business charts and maps. Charts may be displayed within the map or arranged around the map. An algorithm places diagramms such that lines do not intersect and that the overlay of diagrams within a map is reduced to a minimum.

To develop a tool for displaying diagrams, an interface for tabular data sets has been developed. Basically, query results are transfered to the visualization tool via files at the WWW server site. These files contain the tabular data and an description of these, e.g., key lists for identifiers used within the tables. Each column of the tabular data has to be described. Furthermore, interchange formats have to be specified.

The user chooses a visualization goal from a presented goal list, and a parameter to be displayed on the abscissa. Additionally, the layout can be specified; users choose colors, the size of graphics, chart types (line, bar, pie), and the kind to combine them with the map. The graphics may be displayed as fullsize without the map background by a simple mouse click.

3.5 Example

The resulting HTML representation of a combined chart and map service is shown in figure 3. The map contains tabular data for different soil measurements in an industrial region near Heidelberg, Germany. The measurements are visualized as bar charts for different points on the map. The region displayed as a map is limited by the georeferences of the tabular data. The map shows different regions of the hierarchical administrative units: The line represents the town-area, different colors of the background map refer to the districts of the town. The site symbol of an airport is also included in the representation. Above the image, the symbols for the different operations are visible, like zoom or object information. After a symbol has been selected, the action is performed by clicking to a point of interest in the map.

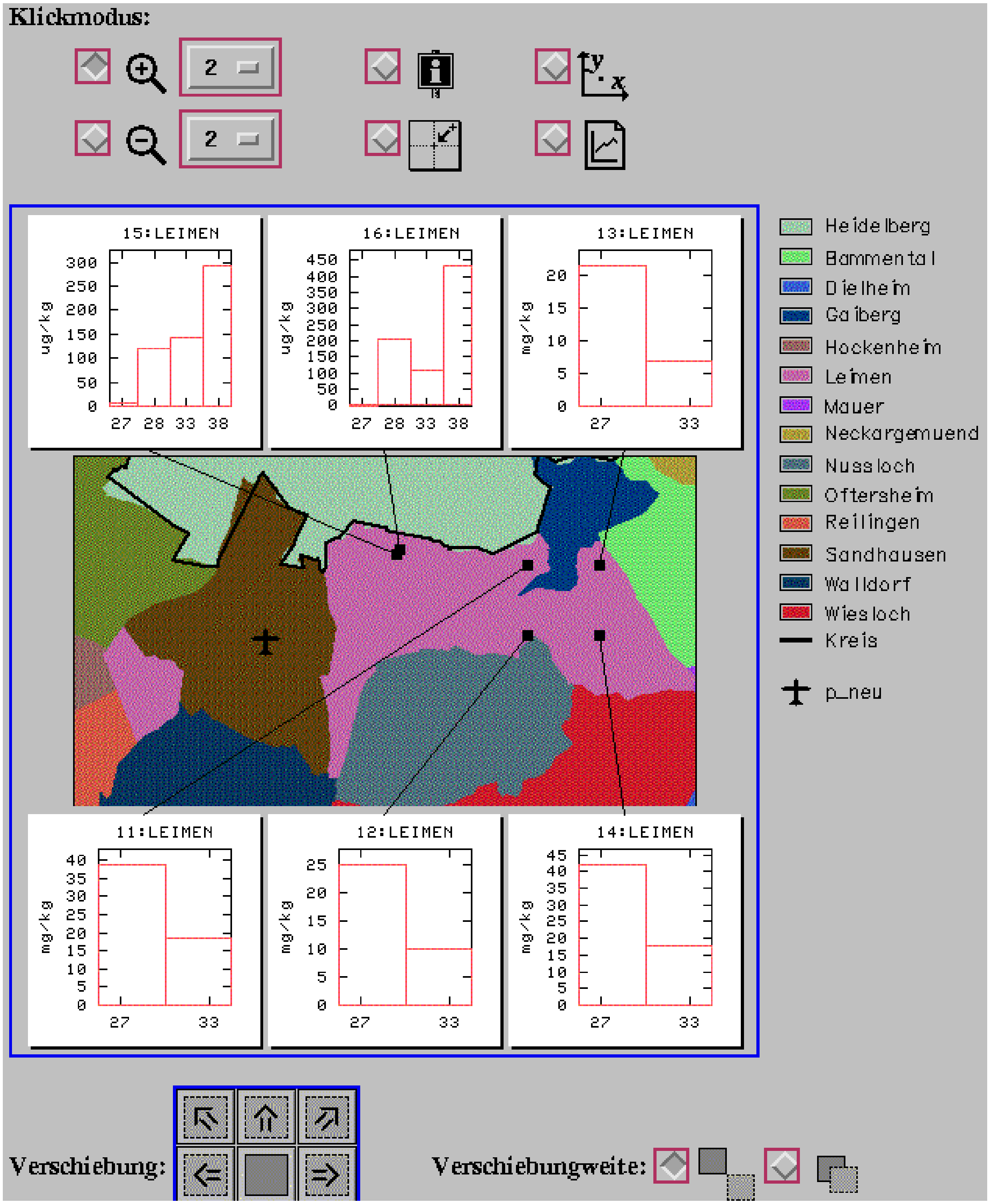


Figure 3: Example of GIS Representation in the World Wide Web

4 A CORBA-based Middleware Layer

4.1 Motivation

WWW makes distributed information sources available. However, pure WWW technology does not support the combination of information from different distrubuted sources in a single HTML page. Information requests (queries) that access different databases or even combine a GIS (at location A) and a database access (at location B) are an example.

The introduction of a middleware layer is a solution to this problem. This layer is responsible for connections between different information sources (servers). Hence, data is still stored and managed at the place where it (organizational) belongs to. An example for data residing at different sources is UDK data [Lessing 96], e.g., one UDK residing in Austria and another one in Germany, each of them containing the country specific metadata. A query should now combine the results from both UDKs, giving a single result page. By using this concept, a federation of servers is built, as shown in the architecture of figure 1.

In order to build such federations of information sources and also to cope with the technical heterogeneity often found in reality, emerging technology standards can be used. An appropriate standard is the Common Object Request Broker Architecture (CORBA), which is described in the next Section 4.2.

4.2 Introducing CORBA

The Common Object Request Broker Architecture (CORBA) is a middleware layer that supports wrapping of distributed object providers (servers) and the use of these servers by clients [Object Management Group 95]. CORBA is defined by the Object Management Group (OMG), which now includes more than 570 members. In this architecture, which is defined independent from any hardware or software, a so-called object request broker (ORB) is responsible for distributing object calls made by clients and server responses to such calls. Important parts of the CORBA specification are a core object model, localization transparency (clients and server need not be aware of their respective locations, e.g. at different hosts), and programming language independence, realized by providing a specific interface definition language (IDL) as well as a dynamic invocation interface (DII) to objects. IDL is mapped to programming language specific bindings, e.g., for C and C++.

Currently, some 15 vendors offer CORBA implementations for different platforms, operating systems and networks. These range from PC systems or Macs using OS/2, Windows or MacOS, to workstations running UNIX and also mainframe environments.

4.3 Incorporating CORBA in the Federation Architecture

As shown in figure 1, the overall software architecture uses CORBA as the middleware layer, into which different system level services are integrated. Many of these system level services (e.g., UDK-based information services, data services for limit values and measurement data) use a relational database system (RDBMS) to store their data. Thus, methods for accessing relational databases are important. We will give an example for this in Section 4.4.

Having developed a CORBA-based architecture, we are able to hide implementation details of services. In principle, these services can be used by any client application. Depending on the final graphical user interface, a client-specific preparation of results is necessary. In our case, we have a WWW server, which operates as a CORBA client. Thus, we provide WWW access to our system level services. Hence, results of system level services have to be prepared as HTML pages. We provide both a service-specific as well as a general preparation of these pages, which is beyond the scope of this paper.

4.4 Example

As an example for integrating a system level service using CORBA, we describe one approach for read access to RDBMS. The CORBA implementations ORBeline [Computing 94b,Computing 94a] and Orbix [IONA 93] have been used. The integration we developed works as follows. A CORBA server process, written in C++, has an IDL interface consisting of operations, which a client uses as database services. The operations in the server process call external C functions that access the RDBMS, e.g., Oracle. In particular, we access the data by using embedded SQL queries that are specified in the C functions. These C functions return a data structure to the server containing the results of the query. We use a linked list of structs as a fairly general though query specific structure for passing query results. In IDL, this linked list is defined as the CORBA predefined aggregation structure `sequence`, giving a kind of one-dimensional array of structs. The server is responsible for encapsulating this data and for passing the results back to the client for further processing. Figure 4 illustrates this process.

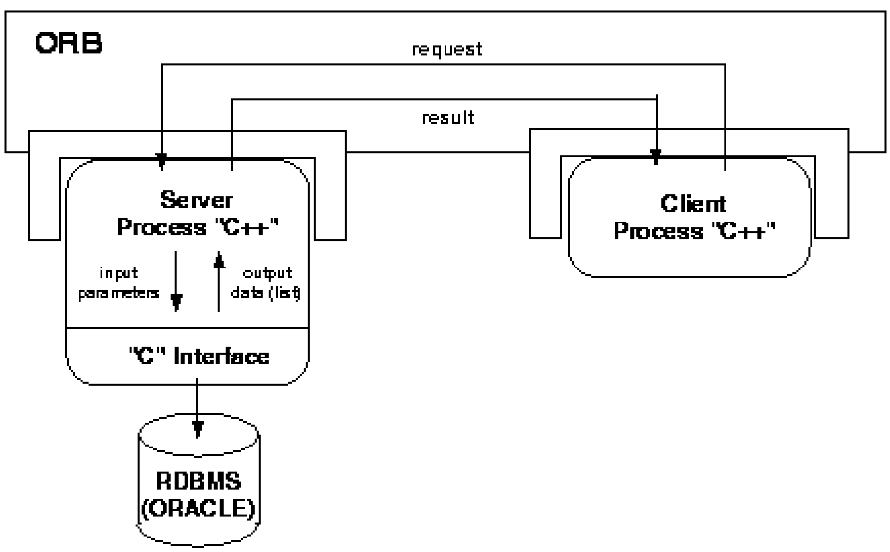


Figure 4: Integrating a relational DBMS into CORBA

The structure information found in the IDL interface file reveals the structure of the data returned by the server. Based on this description, the client is able to manipulate the data as needed. He is responsible for deleting the `sequence` when it is no longer needed.

5 Related Work

In this section, we place our work in the context of GIS and the World Wide Web (Section 5.1) and accessing relational database systems from CORBA (Section 5.2). The list of approaches, concepts, and systems discussed does not claim to encompass all relevant work in the area. We restrict the discussion to those that we think to bear the most direct relationships to our problem and solution.

5.1 GIS and WWW

In a related project [Henning 96] ArcView is used at the client site to display maps. This method has the advantage to avoid processing of data at the server side, but additional software is necessary for every client. Our approach aims at a system which is accessible with a standard WWW browser. This enables the system to provide data and methods to everyone and avoid a lot of cost for large institutions using the data.

Some other scientists are working with GIS and WWW, most of them are situated in the United States. Often informations about the systems are only available on the web itself. Good starting points for a search are [U.S. Census Bureau 95a], [Thoen 95], [Behrens 95] and [OGIS 95]. Some examples of the capabilities of other systems are located at [U.S. Census Bureau 95b], [Huse 95], [Illinois Sate Museum 95] and [ERIN 95]. These systems offer different subsets of the functionality required in EIS. Our approach includes additional features like the combination of maps and graphics, as well as the intercommunication between mapping services and navigation systems as the UDK.

OGIS [Buehler 95] [Buehler (ed.) 94] aims at the definition of an universal, spatial and temporal data and process model. This model may be used for future integration tasks to solve many of the problems caused by heterogeneous GIS. The process of standardization is still going on, so a final solution of integrating GIS objects is not possible currently. Those standards enable an easy access of data and methods in heterogeneous GIS environments, but the style of a final map representation and the kind, datasets can be combined, changes for each individual GIS. In contrast, our approach uses a single GIS for data representation and combination, to enable a unique style and the creation of extended methods for data visualization.

5.2 Accessing RDBMS in CORBA

In several ongoing projects, CORBA is used to integrate RDBMS and other database systems. However, only few projects are reported in the literature. To given an impression of approaches, we refer to some of them.

In the EDS-Navigator-project [Quack 94] a control center for a power plant design and service system has been developed. An approach similar to our one is used to access the database.

For a World Wide customer information system [Hastings 94], a combination of WWW technology and CORBA has been used. An object wrapper layer has been developed, which passes information by pairs of name and value from HTML-forms to method calls of objects, as a generic approach to wrap information sources.

Jupiter [Grimson 94] and MIND [Dogac 95] are prototypes of multidatabase implementations that use CORBA as an interoperability layer. Such systems have to cope with arbitrary (SQL) queries. Thus, a generic database system access-method had to be taken. This method uses generic access query interfaces of the underlying database systems, like a call-level interface, dynamic SQL or a SQL interpreter.

6 Conclusions and Future Work

In this paper, we have presented a federation architecture that provides the functionality of Geographic Information Systems using off-the-shelf WWW browsers. A description of further components of this system currently under development, e. g. online-reports about dangerous waste from the past, keyword-tools, in-depth evaluation of several CORBA implementations and Mosaic's common client interface (CCI) is to be found in [Henning 96].

No additional software tools at the client site are required. Furthermore, the systems supports access to relational database systems. Complex geometric GIS data and tabular data are visualized in maps combined with business graphics. CORBA is used to overcome some of the current limitations of the World Wide Web. The system is currently under evaluation in a private network; hence, we cannot provide an URL here.

Among the points to be addressed in the near future are the transfer of simple tasks from the server site to the client site in order to improve performance, and the integration of heterogeneous, distributed GIS methods and data. The first point will be addressed using JAVA. This new programming language offers a convenient way to transfer small application programs, so-called applets, to the client site. These applets will improve performance with respect to user interactions. Furthermore, they will reduce the transfer data volume by transfering vector data, which then will be rastered for the visualization at the client site. The second point is addressed using a CORBA-based Middleware as described in Section 4. With respect to CORBA and GIS, we carefully take into account the OGIS activities [Buehler (ed.) 94].

Acknowledgements:

The students S. Behrens, U. Beutler, C. Habeck, P. Kirsch, B. Mayer, P. Palmer, C. Rolker, O. Scheffczyk, M. Schoel, and M. Walz were instrumental in the implementation effort at FZI. At IPF, the students M. Böse, C. Hofmann, and H. Krassin were involved in implementing GIS services and visualization tools. We gratefully acknowledge their support.

References

Behrens 95
C. Behrens, C. S. Charles. Geosight project. Technical report, 1995. http://usdac2.rutgers.edu.

Boutell 95
T. Boutell. Gd-library manual and sources. Technical report, 1995. ftp://isis.cshl.org/pub/gd.

Buehler (ed.) 94
K. Buehler (ed.). The Open Geodata Interoperability Specification. In *Draft base Document - OGIS Project Document 94-025R2*, Dec. 1994.

Buehler 95
K. Buehler. The OGIS Technical Committee: Purpose, Status, Future. *Geo Info Systems*, page 48, Jan. 1995.

Computing 94a
PostModern Computing. *ORBeline Reference Manual*. PostModern Computing Technologies, Inc., 1897 Landings Drive, Mountain View, California 94043 U.S.A, 1994.

Computing 94b
PostModern Computing. *ORBeline User's Guide*. PostModern Computing Technologies, Inc., 1897 Landings Drive, Mountain View, California 94043 U.S.A, 1994.

Dogac 95
Ebru Kilic, Gokhan Ozhan, Cevdet Dengi, Nihan Kesim, Pinar Koksal, Asuman Dogac. Experiences in Using CORBA for a Multidatabase Implementation. In *Proceedings of the 6th Int. Workshop and Conference on Database and Expertsystems Applications DEXA'95, London, UK*, September 1995.

ERIN 95
ERIN. Erin australia. Technical report, 1995. http://kaos.erin.gov.au/.

Gleeson 95
M. Gleeson. Fly-tool manual and sources. Technical report, 1995. ftp://www.unimelb.edu.au/pub/cwis/tools/unix.

GNU 95
GNU. GNUPLOT Manual and Sources. Technical report, 1995. ftp://ftp.dartmouth.edu/pub/gnuplot.

GRASS 95
GRASS. GRASS users mailing list. Technical report, 1995. grassu-request@moon.cecer.army.mil, ``HELP" within the mail body.

Grimson 94
John Murphy, Jane Grimson. The Jupiter System: An Environment for Multidatabase Interoperability. Technical report, Dublin City University and Trinity College Dublin, Jan. 1994.

Hastings 94
E. E. Hastings, D. H. Kumar. Providing Customers Information Using the WEB and CORBA. In *2'nd WWW-Conference*, 1994. http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/DDay/hastings/hastings.html.

Henning 96
I. Henning, R. Mayer-Föll, M. Müller, E. Schmid, H. Spandl, A. Koschel, R. Kramer, W. F. Riekert, G. Wiest, W. Geiger, A. Jaeschke, R. Weidemann, F. Schmidt, J. Wiesel.

Projekt GLOBUS - Konzeption und prototypische Realisierung einer aktiven Auskunftskomponente für globale Umwelt-Sachdaten im Umweltinformationssystem Baden-Württemberg, Phase II - 1995, Umweltministerium Baden-Württemberg. 1996.

Huse 95
S. Huse. GrassLinks (A system based on the PD GIS GRASS). Technical report, 1995. http://www.regis.berkeley.edu:80/grasslinks/.

Illinois Sate Museum 95
Illinois Sate Museum. Faunmap (based on ARC/INFO data). Technical report, 1995. http://www.museum.state.il.us/research/faunmap.

IONA 93
IONA. Orbix -- a technical overview. Technical Report PR-TEC-7-5, IONA Technologies Ltd., July 1993.

Kramer 95
R. Kramer, H. Spandl. Metadatenzugriff in Weitverkehrsnetzen: Eine Realisierung am Beispiel des Umweltdatenkatalogs UDK. In F. Huber-Wäschle, H. Schauer, P. Widmayer, editors, *Herausforderungen eines globalen Informationsverbundes für die Informatik; 25. GI-Jahrestagung und 13. Schweizer Informatikertag, Zürich, 18.-20. September 1995/GISI 95*, Informatik Aktuell, pages 610--617. Springer, 1995.

Kramer 96
R. Kramer, T. Quellenberg. Global Access to Environmental Information. In R. Denzer, D. Russel, G. Schimak, editors, *Environmental Software Systems; Proceedings of the International Symposium on Environmental Software Systems, 1995*, International Federation for Information Processing (IFIP), pages 209 -- 218, London, 1996. Chapman and Hall.

Lessing 96
H. Lessing, W. Swoboda, O. Günther. UDK: A European Environmental Data Catalogue. In *Third International Conference/Workshop Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, Jan. 1996. WWW and CD, Santa Barbara: National Center for Geographic Information and Analysis.

McCauley 96
J.D. McCauley, K.C.S. Navulur, B.A. Engel, R. Srinivasan. Serving GIS Data Through the World Wide Web. In *Third International Conference/Workshop Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, Jan. 1996. WWW and CD, Santa Barbara: National Center for Geographic Information and Analysis.

McCool 94
Rob McCool. Common gateway interface overview. Technical report, National Center for Supercomputiong Applications (NCSA), http://hoohoo.ncsa.uiuc.edu/cgi/overview.html, 1994.

Object Management Group 95
Object Management Group. *The Common Object Request Broker: Architecture and Specification*. Object Management Group, Inc., 2.0 edition, July 1995.

OGIS 95
OGIS. Open GIS Consortium. Technical report, 1995. http://www.ogis.org/ogis.html.

Ousterhout 94
John K. Ousterhout. *Tcl and the Tk toolkit*. Addison-Wesley, Reading, MA, 1994.

PBMPLUS 95
PBMPLUS. Pbmplus manuals and sources. Technical report, 1995. ftp://sunsite.unc.edu:/pub/Linux/apps/graphics/convert.

Quack 94
K. Quack. Corba-Implementierung macht Einzelanwendungen zum System. *Computerwoche*, Nov. 1994.

Rodcay (ed.) 95
G. K. Rodcay (ed.). *GIS World Sourcebook 1996*. 1995.

Thoen 95
B. Thoen. WEB - GIS. Technical report, Oct. 1995. http://www.gisnet.com/gis/webgis.html.

U.S. Army Corps of Engineers 95
U.S. Army Corps of Engineers. GRASS Manuals and Sources. Technical report, 1995. ftp://moon.cecer.army.mil/pub/grass/grass4.1.

U.S. Census Bureau 95a
U.S. Census Bureau. Tiger Mapping Service TMS. Technical report, 1995. http://wings.buffalo.edu/geoweb/related.html.

U.S. Census Bureau 95b
U.S. Census Bureau. TMS (based on TIGER data). Technical report, 1995. http://tiger.census.gov/.

Author Information

Arne Koschel, *Ralf Kramer*, *Ralf Nikolai*
Forschungszentrum Informatik (FZI)
Haid-und-Neu-Str. 10-14, D-76131 Karlsruhe, Germany
email: {koschel, kramer, nikolai}@fzi.de

Wilhelm Hagg, *Joachim Wiesel*, *Heiko Jacobs*
Institut für Photogrammetrie und Fernerkundung (IPF)
Universität Karlsruhe, D-76128 Karlsruhe, Germany
email: {hagg, wiesel, jacobs}@ipf.bau-verm.uni-karlsruhe.de

Fri Dec 22 16:39:26 MET 1995

Developing Internet-based user interfaces for improving spatial data access and usability

Chun Sheng Li, David Bree
Department of Computer Science, University of Manchester
E-mail: { csli | dbree }@cs.man.ac.uk

Adrian Moss, James Petch
Department of Environmental and Geographical Sciences,
Manchester Metropolitan University
E-mail: { a.moss | j.petch }@mmu.ac.uk

Abstract: Recent research findings indicate that a lack of awareness, and problems of accessibility and usability of spatial data are significant bottlenecks to increasing numbers of users and applications. Developing easy to use and widely accessible spatial information systems with user friendly and flexible interfaces has been found useful and effective particularly for inexperienced users. New users need to be aware of data services and be able to easily manipulate data. This paper is concerned with the design and development of user interfaces on WWW as knowledge engineering for improving the accessibility and usability of large and complex spatial data sets.

WWW is increasingly used in geo-spatial research and spatial data services. This networked hypertext environment (over the Internet) provides an ideal platform for the KINDS (Knowledge-based Interface to National Data Sets) initiative. The paper describes the design and development of the KINDS experimental system. A survey of academics in the north west of England examining their background, potential use of data sets, knowledge of computing and information systems and human-computer approaches to spatial data has been carried out. Based on the results a series of interfaces have been built on WWW using interactive maps, free text searching and fill-out forms to enable the user to easily discover information about spatial data services and example applications, and directly interrogate spatial data using ESRIs' Arc/Info GIS over the Internet.

Keywords: Spatial data access, digital data libraries, user interface design, information storage and retrieval, WWW, Arc/Info

1. Introduction

The Manchester Information Data sets and Associated Services (MIDAS) at Manchester University provides on-line access to key strategic research and teaching and data sets. These include the UK 1981 and 1991 Census data, Ordnance Survey and Bartholomew mapping data, Landsat and SPOT satellite imagery data. Together these key data sets are known as the 'National Data Sets' (NDS). Within the UK the NDS are freely available to the UK academic community under the Combined Higher Educational Software Team (CHEST) agreement. However uptake of spatial data sets has been found to be limited (MIDAS Annual Report, 1995).

Accessibility and usability of spatial data sets are major bottlenecks to increasing the number of applications (Li *et al.*, 1995). To increase the accessibility of data, data providers must promote awareness of the existence and contents of those data sets. Many spatial data sets are of potential use to a range of different end users however if potential users are unaware of the existence of a data set the result maybe low utilisation and even a waste of an expensive resource (Cornelius and Strachan, 1989; Ruggles 1990; Walker *et al*., 1992). The provision of effective meta data (data about data) underpins attempts to increase data accessibility (McLaughlin and Nichols, 1994). The development and deployment of meta information systems is currently receiving global research attention. See for example GENIE (Walker *et al*., 1992), GeoWeb (Plewe, 1994) and Project Alexandria (Frew *et al*., 1995; Andresen, D. *et al*., 1995).

Poor usability results from low understanding of the use of spatial data. Spatial data analysis is a knowledge intensive activity. New users face a significant learning curve when adopting spatial data. The techniques associated are often significantly different from other types of analysis (Fotheringham and Rogerson, 1993). Users must familiarise themselves with the command structure and nomenclature of geographical information systems. Often users must also have expertise in data pre-processing and the systems of data provision used by data providers. Often data sets have to be used together with ancillary software packages to support operations such as data handling, map plotting, etc. To be able to use the data, the user should have both knowledge of data coverage, structure and the support systems. The combinations of high level technical skills required often result in spatial data handling being the preserve of the highly technically competent.

The Knowledge-based Interface to National Data Sets (KINDS) project aims to extend and intensify the use of the National Data Sets service available from MIDAS by educating potential users in their use. This is a dual mission. Firstly potential users are to be made aware of the existence of the data sets and secondly guided in their use. In order to achieve this, KINDS must reach the widest possible audience through the World Wide Web (WWW) with user friendly, easy to access and effective Internet-based search engines and hypertext interfaces for users to browse and handle spatial data.

This paper presents some selected findings from two surveys of potential users examining actual and potential use of the NDS; users' knowledge of computing and information systems and human-computer approaches to spatial data. The survey findings have been used as the basis for the development of KINDS experimental system. The paper describes the design and development of a series of interfaces on WWW to enable the user to easily discover information about spatial data services and example applications, and carry out real-time task processing with ESRI Arc/Info to generate maps over the Internet.

2. KINDS technical survey and findings

An extensive user survey including both semi-structured interviews and a technical questionnaire survey was carried out in Manchester at the early stage of the KINDS project to examine the academic requirements for data in teaching and research, and approaches likely to be used to accessing spatial data.

The Manchester academic community is amongst the largest in Europe. Seventy eight researchers, lecturers and post-graduate students at academic departments representing about 28 disciplines were approached to determine their needs for spatial data. The rationale of the user survey was to reveal the extent of data use within the academic community and paths of dissemination through users. Hence a mixed approach of quantitative and qualitative research methods was adopted. One of the prime objectives of the survey was to examine informal methods that users employed to gain support. This is the phenomena of a user who when faced with a problem using a computer system opts to speak to a nearby college or friend rather than approaching formal support services. In order to test the extent to which this took place in the data processing environment respondents were asked to suggest likely colleges to be interviewed. Whilst we fully acknowledge that this process does not adhere to random sampling rules and precludes rigorous statistical analysis a fuller picture of information dissemination within higher education was revealed as a result. The full results of the user survey will be reported in a future publication. We present some of our analysis here as background to what will follow.

3*2.1 Actual and potential use of the data*

Use of spatial data is a knowledge intensive activity. Spatial analysis concepts and techniques are significantly different from those used in other analysis methods (Fotheringham and Rogerson, 1993). Users must familiarise themselves with the command structure and nomenclature of geographical information systems and the environment in which they reside - often GIS are mounted on high end workstations whilst the majority of new users are familiar with personal computers. The actual use of MIDAS data services appeared relatively low (see left columns in Figure 1). However many respondents felt that national data sets would be useful resources when they became aware of the coverage, contents and suitability of the data sets. The right columns in Figure 1 show the users' interest in using national data sets for their teaching and research.

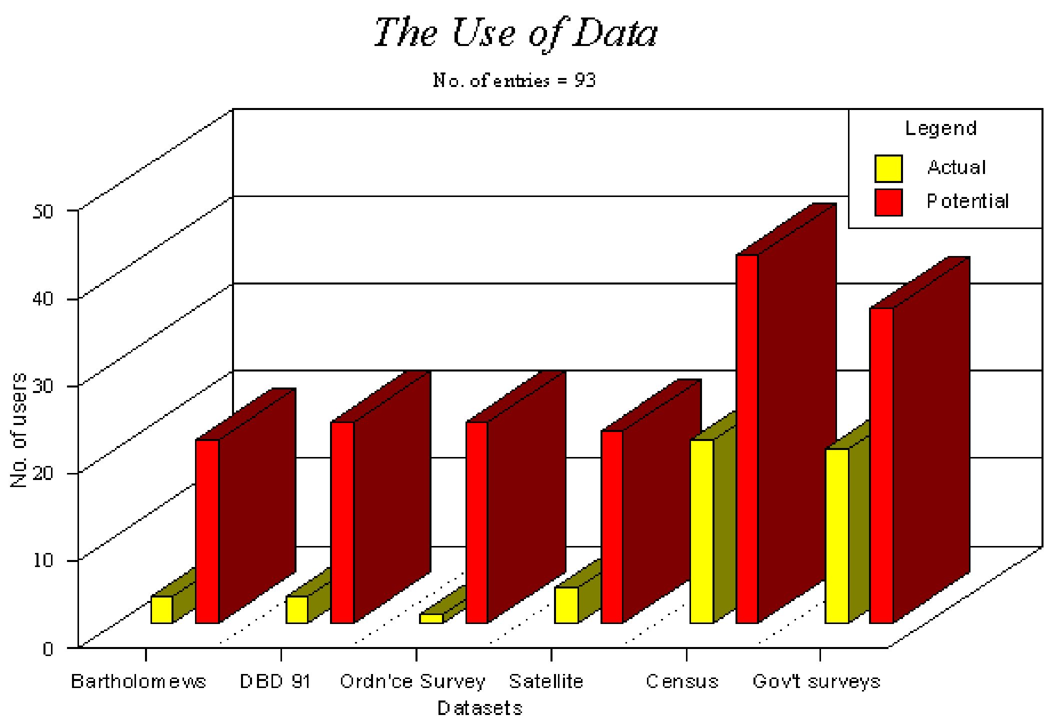


Figure 1. Actual use vs. potential use of data

2.2 The user's technical ability

Knowledge of information technology seems to be a key factor which decides whether data set applications are successful. To investigate how these users access data and how user friendly and effective interfaces are for spatial data handling, a technical questionnaire survey was carried out. User knowledge, interest in networked information systems and human computer interaction issues were examined.

41 (52.5%) respondents returned completed questionnaires. The users were classified into four categories based on knowledge and daily use of information systems and computer-based applications.

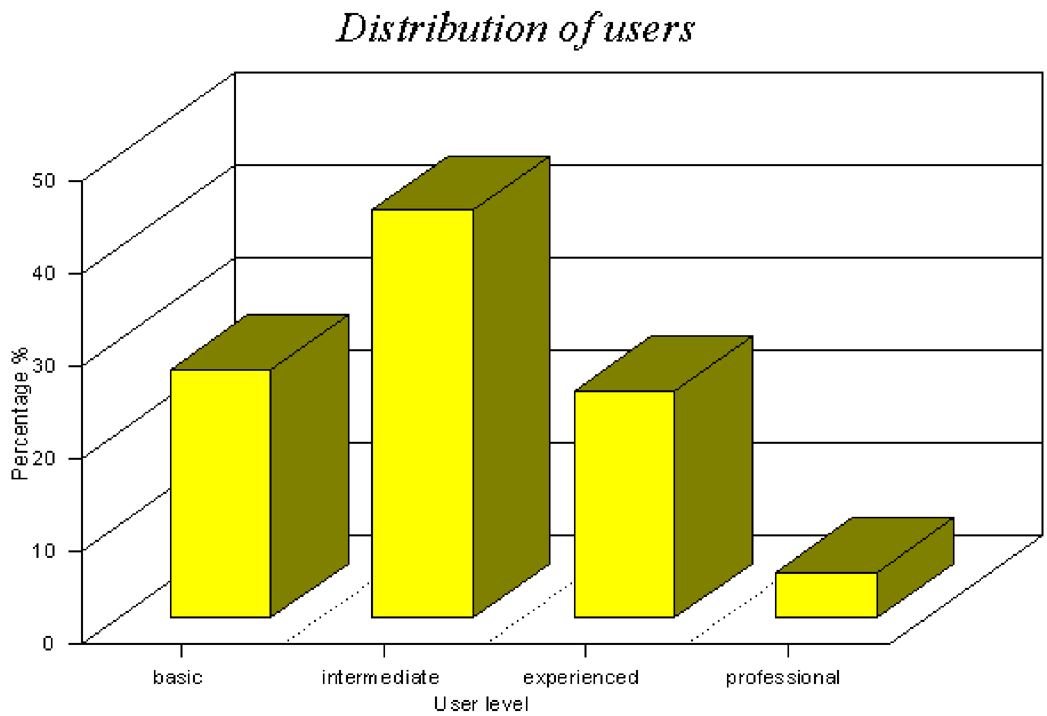


Figure 2. Categories of the users

The extensive use of the Email facility indicated that most users already have access to computer networks. The Internet utilities (telnet, FTP, etc) were also widely used. WWW, the most recent Internet service, has become one of major platforms in the academic community. Over 70% respondents commented that WWW has been a useful resource for their teaching and research. The following table shows usage of the Internet-based Information systems against each category of user.

LEVEL - user level by NETINFO - usage of networked information systems

LEVEL / Count	NETINFO: news group 1	gopher, wais 2	WWW 4	WWW + others 6	Row Total
1 basic		1	3	2	6 17.1
2 intermediate	5	1	1	10	17 48.6
3 experienced				10	10 28.6
4 professional				2	2 5.7
Column Total	5 14.3	2 5.7	4 11.4	24 68.6	35 100.0

Number of Missing Observations: 6

Table 1. The Internet-based information systems being used

2.3 User interfaces

The user interface is one of the most important aspects of a spatial information system. Respondents were asked to select the types of interfaces they prefered. Selections were listed in sequence according to the respondents preference. The results were processed using weighting and ranking techniques, to assess which class of users found which interfaces more useful. Table 2 lists the interfaces ranked by the points scored.

Variable	Std Dev	Minimum	Maximum	Sum	Valid No	Label
A	1.24	2	6	184.00	36	interactive dialog boxes
B	1.19	2	6	124.00	29	tree structured menus
D	1.28	2	6	121.00	27	interactive maps
C	1.15	3	6	102.00	24	multi-selectable items
E	1.35	2	5	71.00	23	natural language interfaces

Table 2. Answers ranked by the total points scored

Some users were unable to comment on interfaces since they had no experience. To address this data on the respondents' background was considered. The following table shows the ranking indicated by the average points scored.

Variable	Mean	Std Dev	Minimum	Maximum	Valid No	Label
A	5.11	1.24	2	6	36	interactive dialog boxes
D	4.48	1.28	2	6	27	interactive maps
B	4.28	1.19	2	6	29	tree structured menus
C	4.25	1.15	3	6	24	multi-selectable items
E	3.09	1.35	2	5	23	natural language interfaces

Table 3. Answers ranked by the average points

2.4 Bottlenecks for the spatial data use

The survey revealed that many respondents were unaware of the spatial data service and the availability of data sets from MIDAS. Factors such as the significant spatial data handling learning curve, inadequate technical guidance and support help to explain lack of use.

Ease of use of user interfaces was preferred over functionality by users with less technical skill. The results of the survey were broadly in compliance with Davis and Medyckyj-Scott (1994) who reported that inexperienced users suffered difficulty transferring existing knowledge from their disciplines to spatial data handling. The results indicate that user friendly and flexible interfaces are important for improving spatial data usability.

3. Browsing spatial data

3.1 Bartholomew Map Data set

The Bartholomew data has been used as the test data by the KINDS project. It is a layered vector map data set comprising of point, line and area features. The data is structured into several sets comprising World, European, GB, London and Central London coverage's. In the Bartholomew (Great Britain) data set, the coverage is divided into tiles, based on the National Grid. Each tile, covering an area of 100 km square, is identified by a set of two characters as shown in Figure 3.

The data for either the GB national coverage or a tile is stored in 16 thematic data layers including administrative boundaries, contours, roads, railways and ferries, point features, urban areas, and water, etc. In addition, there is an annotations layer that contains textual information describing the feature data.

Within the Bartholomew data set, the features are organised into classes, each class is identified by an OBS_ACC_NO (observation accession number), which describes the feature and its entity types (point, line or polygon). The OBS_ACC_NOs uniquely identify each type of feature. For instance in the 'Roads' data layer, existing motorways are the unique code no (235), the primary trunk dual carriage way A roads are referenced as (173291) and so forth.

The following is a list of thematic data layers available in the Bartholomew (GB) data,

Data Layer	Description
Administrative Boundaries	National and Regional boundaries. Also includes lochs(lakes) and coastline.
Contours	Contours at 100m intervals plus 50m and 150m.
Danger Zones	
Drainage	Also includes canals.
Forest Parks	
National Parks	
National Trust	
Other Lines	Includes other rights of way.
Points	Leisure, physical, road and industrial features and other transport and road distance points.
Roads	Motorways, A and B roads (Dual or single carriageway), minor roads and roads under construction.
Railways & Ferries	
Regional Parks	
Scenic Areas	
Topography	Rocky shores, beaches and woodland.
Water	Lochs(lakes) and marshes.
Urban Areas	
Annotation	Cartographically placed text annotation.

3.2 The KINDS thematic map library

Potential spatial data users can gain information about the data set far more easily by browsing its contents than by reading a textual description. Thus enabling users to quickly browse through maps was a major objective in increasing awareness of the Bartholomew map data. The WWW is a fast and feasible way of presenting maps by distributing GIF format images. Such images which are of sufficiently good quality to display spatial objects but require only small amounts of memory and transmission time through the internet to the user's client software.

A number of sample maps were generated manually using ESRI Arc/Info Arcplot in the early stages of the KINDS project as a feasibility study. After the KINDS experimental system was released on WWW, users expressed an interest in seeing more detailed feature maps covering more specific areas. A feature map library of the major features contained in the Bartholomew (GB) map data was built to provide as detailed spatial information as possible. Spatial features are linked to their corresponding textual descriptions in hyper text markup language (HTML) pages. the features and their descriptions can be retrieved by either an interactive map interface or a search engine.

A virtual map library has been created. The library comprises of a full UK national coverage directory and 55 (for all 100km squares) sub-directories named after the corresponding tile of the National Grid. ESRI Arc/Info Arc Macro Language (AML) was used to create scripts to automatically generate feature maps from the data set and generate legends automatically. Arcplot is unable to export maps in GIF format and so SDSC Image Tools (produced by San Diego Supercomputing Center) was used for converting maps from Sun raster to GIF format.

3.3 Map interface

An interactive map (also sometimes referred to as a clickable image) is an inline image in an HTML document. The position of any mouse click within the image is captured using of the HTML tag ISMAP. When a user clicks the mouse over the image, the browser sends the pixel co-ordinates to the WWW server. The co-ordinate information is then processed by a program on the server to return an appropriate URL (HTML document) after comparing the mouse co-ordinates with with boundary location information in the virtual library imagemap database. The ISMAP tag provides for a limited degree of spatial querying of maps.

The required number of co-ordinates is dependant upon the shape of region to be defined. For circles, two pairs of co-ordinates are required: centre and any edgepoint; for rectangles, co-ordinates of upper-left and lower-right; for polygons of 100 vertices at most, each co-ordinate pair stands for a vertex.

The ISMAP facility is a powerful tool for producing WYSIWYG - "what you see is what you get" interfaces. The interactive map provides an intuitive and easy to use method for the user to swiftly browse through the Bartholomew data's "layer-and- tile" structure. The interactive map is based on a UK map using the National Grid, regional and county boundaries for geographical location referencing (Figure 3). Countries are marked in different colours for users to easily point to an exact area of interest, for example, Scotland is in blue colour, England in red and Wales in green. The user can simply move the mouse to an area (tile), and click on it to see detailed thematic data which is linked to the KINDS map library with over 800 feature maps.

3.4 Free text search engine

A "free text" search engine provides a direct entry for both expert and inexperienced users to quickly discover information about the data set. The interface uses a dialogue box for the user to enter a query which is passed to the search engine. The engine then retrieves an index file linking to the KINDS thematic map library and returns headings of documents which match the users query.

The index file is an important element in the free text search engine. Spatial features originally organised based on National Grid references have been restructured to be linked with UK counties/regions, and major cities. Each entry in the file is indexed with both the two-character identification of tile and counties or regions as well as major cities covered by the tile. Thus the search engine complements the map interface by allowing the user to search for information about specific cities rather than using the technical tile structure.

The search engine executes a Perl script in the server. It filters out terms (with less than two characters or starting with symbols such as "/", ".", "*") which may cause inaccurate results to be presented. Boolean (logic AND, OR) searches and right-hand truncation searches (with an "*" at the end of search term) are supported to make searches effective and efficient.

Help information including a list of indexed terms used in the search engine and information about its structure have been added for inexperienced users. The user is also able to start querying by clicking on 'hyperlinks' listed in the help file.

4. Making Road Maps - an example for directly handling spatial data across WWW

4.1 Representation of UK Roads

A dynamic link between the WWW and the Bartholmew data using ERSI Arc/Info has been created to familiarise the user with working with spatial data. The Bartholomew (GB) data set is structured into 16 thematic data layers comprising over a thousand classes of features, the roads data layer was selected for the dynamic link experiment. The roads layer includes 22 classes of road features (see the list below) and each feature is available in both the national coverage and individual tiles.

OBS_ACC_NO	Description	Feature Type
235	Motorway	Line
245	Motorway Under Construction	Line
173287	Motorway Tunnel	Line
173291	A Road Primary Trunk Dual C/W	Line
173292	A Road Primary Trunk Single C/W	Line
173293	A Road Primary Trunk Passing Places	Line
233	A Road Primary Non-Trunk Dual C/W	Line
226	A Road Primary Non-Trunk Single C/W	Line
173294	A Road Prim. Non-Trunk Passing Places	Line
231	A Road Non-Primary Dual C/W	Line
227	A Road Non-Primary Single C/W	Line
222	A Road Non-Primary Passing Place	Line
243	A Road Dual C/W Under Cons (all)	Line
241	A Road Single C/W Under Cons (all)	Line
232	B Road Dual Carriage Way	Line
230	B Road Single Carriage Way	Line
229	B Road with Passing Places	Line
244	B Road Dual C/W Under Cons (all)	Line
242	B Road Single C/W Under Cons (all)	Line
73	Road Tunnel	Line
130	Minor Road	Line
173295	Private Road	Line

The UK roads are well classified and structured in the data set. The following taxonomic conceptual model shows the hierarchical relations of roads.

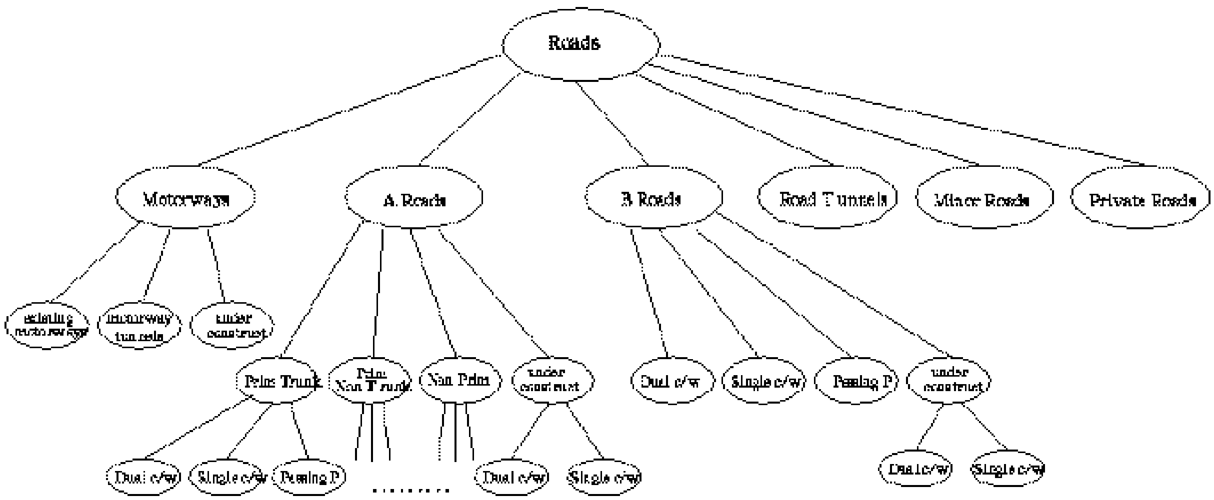


Figure 4. The hierarchical structure of roads

The Bartholomew data is mounted on the MIDAS national data sets machine as Arc/Info coverages. To assess the suitability of the data set, users must first manipulate it within Arc/Info. Therefore users must already have considerable knowledge of spatial data handling before being able to make a decision about the use of MIDAS data. New users are thus placed at a considerable disadvantage. The KINDS experimental system reverses this situation by allowing users to manipulate the data set without having to directly use a geographical information system. A visual approach based on a clickable UK map (as in Figure 5) is adopted for choosing the coverage. The user can select features they wish to see using a simple forms interface.

4.2 WWW form-based interface - *a front end to the map maker*

WWW forms (or fill-out forms) are a computer equivalent of paper forms. When users 'submit' a WWW form the users responses are transmitted to a program on the HTTP server for processing.

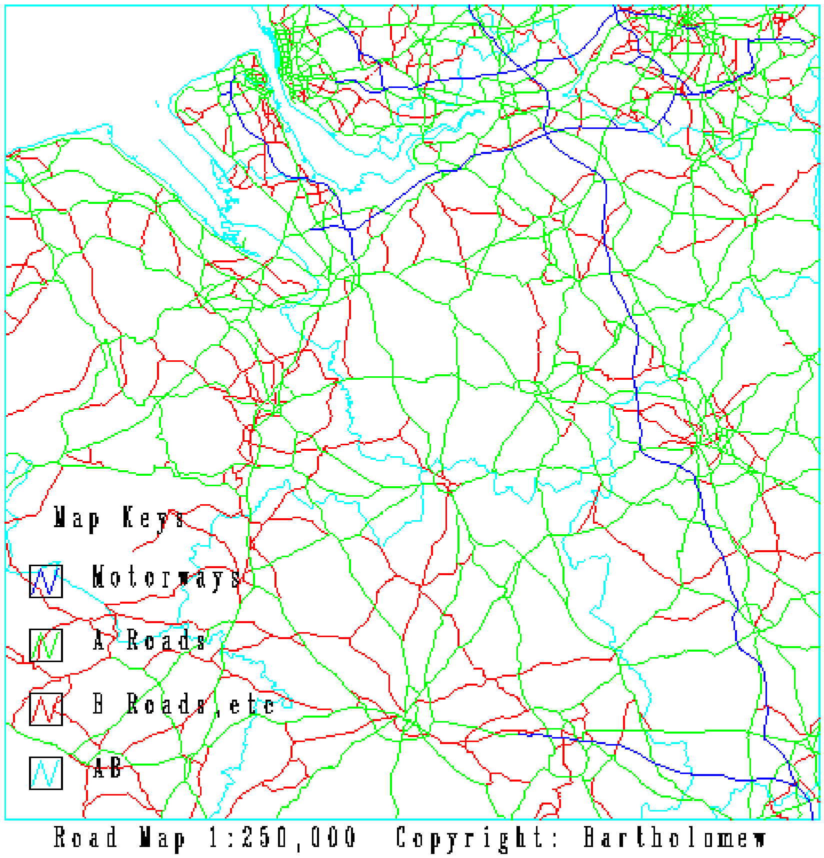
A form provides for complex interactions between the user and other software via programs residing in the HTTP cgi-bin (executable programs) directory.

A forms interface to the ESRI Arc/Info system has been created to allow WWW users to interact with spatial data (Figure 5). The interfaces allows the user to create simple maps which inform about the contents and potential of the Bartholomew data set. The test application allows the user to select and display elements of the road data layer of the Bartholomew data set.

4.3 AML scripts coding - *an automatic process*

Submitting the WWW form activates Arc/Info after generating an AML script via a translation program residing in the HTTP servers' common gateway interface (CGI). The translation program is implemented using Perl. The query string is split into items to determine where to access the data (i.e. which tile); what road features and additional themes are selected and what colours should be used to mark these features.

Subsequently an AML script then can be coded to produce a map reflecting the user's request. A file comprising map symbols and textual descriptions for generating map keys (legends) with the AML script is also produced. The result of the AML script is a 7" by 9" road map in encapsulated post script (EPS) format. The EPS file is then converted into GIF format using the Image Tools package. The process in its entirety normally takes about 60-80 seconds to complete and send a map (9k to 21k in size) back to the users WWW browser (this is dependant upon on the speed of the users network). The above map is an example road map of south Manchester and Cheshire generated according to the user request in Figure 5.



Example of a Road Map
(Figure 6)

After the Arc/Info working environment being set up,

4.4 Outline of the KINDS Map Maker
The following schema illustrates the major processes of the KINDS experimental system.

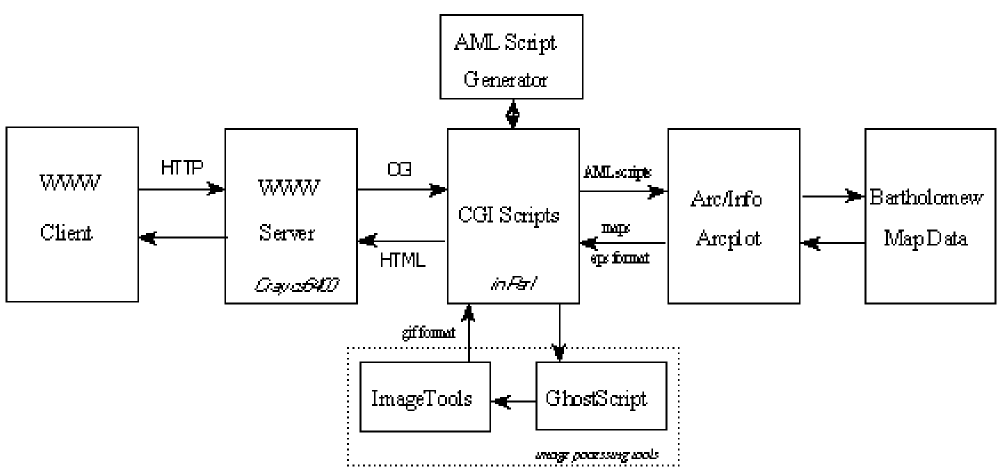


Figure 7. Outline of the KINDS map maker

5. Future development

KINDS is a three year initiative at the end of the first 18 months of funding. In the remainder of the funding period further data sets and software packages will be added to the existing frame work. The existing interface will be complemented by a knowledge base to guide the users through use of the data. Knowledge acquisition for the development of the KBS is underway and coding will commence shortly.

Acknowledgements

Andrew Johnston, of the KINDS team, deserves special mention as co-worker in carrying out the user surveys. In addition we wish to thank the other members of the KINDS team (Keith Cole, Kamie Kitmitto, Jim Yip and Andrew Basden) not least for their help and advice. The Knowledge-based Interface to National Data Sets (KINDS) project is a multi-institution iniative comprising MIDAS (Manchester Computing, UK), the Department of Environmental and Geographical Sciences (Manchester Metropolitan University, UK) and the Information Technology Institute (University of Salford, UK). KINDS is funded by the Joint Information Systems Committee (New Technology Initiative NTI-107).

Bartholomew digital data is Copyright Bartholmew and available to UK academics from MIDAS under the terms of the CHEST license agreement. We wish to thank Dr. Tim Rideout of Bartholomew for his support and helpful comments. Bartholomew may be contacted at Dr. Tim Rideout, 12 Duncan St., Edinburgh, EH9 1TA, UK.

ESRI, Arc/Info, Arcplot and ARC Macro Language are registered trademarks of the Environmental Systems Research Institute, Inc., Redlands CA, USA.

References

Andresen, D., Carver,L., Dolin R., Fischer, C., Frew,J., Goodchild, M., Ibarra, O., Kothuri, R.,
Larsgaard, M., Manjunath, B., Nebert, D., Simpson, J., Smith ,T., Yang, T. and Zheng, Q.(1995)
The WWW prototype of the Alexandria digital library

http://alexandria.sdc.ucsb.edu/public-documents/papers/japan-paper/

Berners-Lee, T (1993) The WWW initiative and HTTP, HTML, etc. CERN WWW Documentation.

http://www.w3.org/

Boston, T. and Stockwell, D. (1994) Interactive species distribution, mapping and modelling
using World Wide Web, Proceedings of the Second International WWW Conference 94, Chicago, USA.

http://kaos.erin.gov.au/database/WWW-Fall94/species_paper.html

Crossley, D and Boston, T (1995) A generic map interface to query geographic information
using the World Wide Web.

http://www.w3.org/pub/Conferences/WWW4/Papers/australia/

Davis, C and Medyckyj-Scott, D (1994) GIS usability: recommendations based on the user's
view. International Journal of Geographical Information Systems, vol.8, no.2, pp175-189.

Frew, J., Carver, L., Fischer, C., Goodchild, M., Larsgaard, M., Smith, T. and Zheng, Q. (1995)
The ALexandria rapid prototype: buidlding a digital library for spatial information

http://www.esri.com/resources/userconf/proc95/to300/p255.html

Fotheringham, A.S. and Rogerson, P.A. (1993) GIS and spatial analytical problems.
International Journal of Geographical Information Systems, vol.7, no.1, pp3-19.

GENIE Project (1994).

http://www-genie.lut.ac.uk/info.html

Li, C S; Kitmitto, K and Cole, K (1995) Developing a WWW interface to Arc/Info. Proceedings of ESRI'95 (UK) user conference, Nottingham University, Sept 14-15, 1995.

Li, C S; Moss, A et al (1995) Access Large and Complex Data sets via WWW. Proceedings of NTTS '95, Nov 20-22, 1995, Bonn, Germany.

Massem, P (1994) Demo of the interface between Arc/Info and the web.

http://www.geo.ed.ac.uk/home/RESEARCH/MASSEM.HTML

McLaughlin, J. and Nichols, S. (1994) Developing a national spatial data infrastructure. Journal of Surveying Engineering ASCE 120 2 pp62-76.

McCool, R (1994) The Common Gateway Interface, NCSA WWW Documentation.

http://hoohoo.ncsa.uiuc.edu/cgi/overview.html

McGranaghan, M (1991) Matching Representations of Geographic Locations. In Mark, D M and Franks A U (eds), Cognitive and Linguistic Aspects of Geographic Space, pp387-402.

Medyckyj-Scott, D and Blades, M (1991) Cognitive representations of space in the design and use of geographical information systems. People and Computer VI, British Computer Society Conference Series, 1991, pp421-434.

Petch, J; et al (1995) The KINDS project. Proceedings of NTTS '95, Nov 20-22, 1995, Bonn, Germany.

Petch, J; Moss, A; Johnston, A and Yip, J (1995) Spatial data services: analysis of user needs. (to appear)

Plewe, B (1994) The GeoWeb project. Proceedings of the Second International WWW Conference 94, Chicago, USA.

http://wings.buffalo.edu/~plewe/paperwww.html

Raper, J. and Green, N. (1992) Teaching the principles of GIS: lessons from the GISTutor project International Journal of Geographical Information Systems vol.6, no.4 pp279-290.

Walker, D.R.F., Newman, I.A., Medyckyj-Scott, D.J. and Ruggles, C.L.N. (1992) A system for identifying data sets for GIS users. International Journal of Geographical Information Systems vol.6, no.6 pp511-527 .

James Darrell McCauley, Kumar C. S. Navulur, Bernard A. Engel, Raghavan Srinivasan

Serving GIS Data Through the World Wide Web

ABSTRACT

The Internet, a world-wide collection of interconnected networks of computers, has facilitated the accessing and sharing of information around the globe. The World-Wide Web is a project on the Internet that allows hypermedia information retrieval across the network. Geographic Information System (GIS) data were made accessible on the Internet by using the Geographic Resources Analysis Support System (GRASS) and the Common Gateway Interface (CGI) of the Hypertext Transfer Protocol in the World-Wide Web. An X Window System-based GUI enabled users anywhere on the Internet to manipulate and display GIS data layers of interest. A platform-independent, display-only map production system was also developed for data browsing. Data were also organized using visual search techniques (image maps) and made available in a vendor-neutral format (Spatial Data Transfer Standard).

KEY WORDS:

Geographical Information Systems (GIS), Internet, World-Wide Web, Data Distribution

INTRODUCTION

Availability of spatial data in digital forms has long been one of the most significant obstacles to the widespread development and use of GIS applications. The first two International Integrating GIS and Environmental Modeling Conferences/Workshops identified digital spatial data availability as one of the greatest problems facing those developing integrated modeling and GIS applications. To overcome this, numerous data development efforts are being planned, are underway, or have been recently completed to provide digital spatial data. As such efforts develop spatial data in digital form, the data availability obstacle becomes a data access obstacle. Several access-related problems also arise including: (1) locating spatial data in digital form, (2) obtaining access to the desired data in a timely fashion, (3) accessing the most recent version of spatial data sets, (4) compatibility of data formats, (5) large storage requirements for data of interest, and (6) exchange of fees for access to data. The information super highway (the Internet) provides an opportunity to overcome many of these problems and to facilitate the application of GIS--even for the novice user.

The overall goal of this work is to demonstrate the potential of Internet for overcoming spatial data access problems and for facilitating the use of GIS by large numbers of diverse users. The Internet is an international communication infrastructure comprised of thousands of regional networks scattered throughout the globe (Comer, 1995). This world-wide connectivity includes more than 13,500 foreign networks and over 20 million users in over 50 countries. Presently, more than 15,500 billion bytes of information are transferred per month across this network. The

Internet supports multiple modes of communication such as electronic mail, remote login sessions, remote file transfers (FTP), hypermedia searches, etc. The Internet is evolving as a major medium for information sharing and retrieval and is an efficient tool for scientific and engineering research and development.

The World-Wide Web (also called WWW, W3, or the Web) is a wide-area hypermedia information retrieval system providing access to a myriad of documents and data on the Internet. WWW is also a body of software and a set of protocols and conventions to provide easy and consistent access to information on the Web. There are many W3 browsers that can be used to surf the Web, including Mosaic (NCSA, 1995), Netscape (Netscape Communications Corporation, 1995) and Lynx (Montulli, 1995).

With the development of the WWW, opportunities arose to organize various resources found on the Internet in an efficient and user-friendly manner. One major type of resource is geographic information systems (GIS) data.

OBJECTIVES

The main objectives of this work were as follows:

- Demonstrate the use of Hypertext Markup Language (HTML) in organizing and sharing available GIS data;
- Provide access to GIS software on the Internet; and
- Explore and demonstrate the capabilities of GIS via Internet

BACKGROUND

First, a brief overview of important protocols and conventions of the WWW is given. These include the Hypertext Markup Language, the Hypertext Transfer Protocol, and the Common Gateway Interface. Following this overview is a brief tour of an integrated GIS data organization, browsing, analysis, and transfer system. Finally, implementation details for this system are given.

Hypertext/Hypermedia

The operation of the Web relies mainly on hypertext and hypermedia as means for interacting with the users. Hypertext is essentially the same as regular text but it `points' to other documents---in the case of hypertext on the Web, these other documents are on the Internet. Hypermedia is hypertext with a slight difference. Hypermedia documents have hyperlinks not only to text but other multimedia forms, such as images, sound files, video files, etc. Hypermedia can be viewed as a combination of hypertext and multimedia. Hypertext Markup Language (HTML) is the standard markup language for creating and recognizing Web documents (Berners-Lee and Connolly, 1995). As a markup language, HTML allows the user to control information presentation in a number of ways (e.g., fonts selection and size) without changing the original content. Authors of HTML documents define the logical structure and control the information content without concerning themselves with text formatting (e.g., fonts and colors). HTML documents are typically 7-bit ASCII files with formatting codes that contain information about document structure and hyperlinks.

WWW uses Uniform Resource Locators (URLs) to represent hypermedia links and links to network services (hyperlinks) within HTML documents (Berners-Lee, 1994b). The general structure of a URL is: protocol://host:port/path where

```
protocol is the Internet protocol (e.g., http, ftp, news),
host is the name of a host computer (in RFC1037 format (Hunt, 1992
     and Albiltz and Liu, 1992) connected to the Internet,
port is an optional integer value specifying a host port, and
path is a filename.
```

Many of the references in this paper contain URLs.

Hypertext Transfer Protocol (HTTP)

Web software is designed around a distributed client-server architecture. A Web browser is client software that can send requests for documents to any Web server. A Web server is a program that, upon a request for a document, processes the client's request and sends back the document or an appropriate message (e.g., an error message). The processing of a request is done by the server and presentation of data is left to the client. The Hypertext Transfer Protocol (HTTP) is the language used by the Web client-server interface (Berners-Lee, 1994a). It utilizes the Transmission Control Protocol/Internet Protocol (TCP/IP) for communication between Internet hosts (Hunt, 1992). This process of client-server interaction involves the following process:

```
Connection:
   The client establishes a TCP/IP connection to the server using the
   URL address, usually using port 80;
Request:
   The client sends a request for the URL using HTTP;
Response:
   The server processes the request and sends back the requested document;
Close:
   The server closes the connection and client terminates the TCP/IP
   connection.
```

All Web clients and servers must be able to speak HTTP in order to send and receive hypermedia documents.

There are many Web servers, each differing slightly in functionality and implementation. These servers, like the FTP daemon (Postel and Reynolds, 1985)are programs that respond to an incoming connection and provide a service to the client. Hypertext Transfer Protocol Daemon (HTTPD) (NCSA, 1995a) is a public domain Web server developed by the National Center for Supercomputing Applications (NCSA). HTTPD also records the date and time of requests along with the IP number of the client, which is useful for keeping track of traffic.

Common Gateway Interface

The Common Gateway Interface (CGI) (McCool, 1995) is an interface under a Web server, such as HTTPD, for running external programs or gateways. CGI facilitates the handling of the information requests and can act as a gateway for returning the appropriate document or creating a document on-the-fly. With a CGI, a Web server can provide information which is not in a form readable by the client (e.g., GIS data files) and can act as a gateway between the server and the client for interaction.

Gateway programs or scripts are server-side executable programs that are run (upon request from a client) to serve information. These gateways are initiated when the client requests the URL corresponding to the gateway. Since these scripts are executed on the server, gateway programs are somewhat independent of the client's operating environment. Gateways interact with the client and server using the HTTP. Gateways conforming to the HTTP specifications can be developed in any programming language, such as C, FORTRAN, Pascal, PERL, Bourne Shell, C Shell, etc.

Information requested from the server to the CGI script is handled using command line arguments as well as environment variables. The environment variables are defined when the server executes the gateway program.

Examples of environment variables are:

REQUEST_METHOD
The method with which the request was made. Request method "POST" is generally used in HTML forms.

QUERY_STRING
Arguments to a CGI program.

REMOTE_ADDR
The IP number of the client making the request. This can be used for defining the remote DISPLAY when running X Window System programs.

AUTH_TYPE
If the server supports user authentication for security reasons, this variable is required for protocol-specific user authentication.

CONTENT_TYPE
This defines the type of data attached with the request to the server.

CONTENT_LENGTH
The length of the content attached to the URL which is required to decode the CONTENT of the request from the client.

CGI scripts can return a variety of document types, such as images, HTML documents, and audio files, in response to the request from the client. Information on the type of document/data that is being sent back depends on the content in the first line of the response. The first line will be different depending on whether the program is returning a full document or a reference to one. In the former case, the first line of the gateway output should be of the form:

Content-type: (a MIME type/subtype encoding (Borenstein and Freed, 1992)).

For an HTML document, the first line of the output is:

Content-type: text/html

Immediately following this is a blank line/linefeed, which indicates to the server that the definition of the output is over.

Sensitive images or image maps are gateway scripts that can be used to make an image region sensitive with hyperlinks pointing to different URLs. The imagemap program is public domain software written in C that provides the above functionality. The image map software requires an ASCII map file that contains coordinates defining the regions (polygons, rectangles or circles) and the corresponding URL to be fetched. The ASCII map file can be prepared using the software xv or mapedit, which are shareware and public domain software, respectively.

Geographically-sensitive image maps were used in the present research to arrange spatial data sets.

RESULTS AND IMPLEMENTATION

Many elements of an integrated WWW-based GIS data organization, browsing, analysis, and transfer system are discussed below. These elements allow the following interaction:

- Using image maps based on political or natural (e.g., watershed) boundaries, clients may visually search, locate, and select data of interest.
- Maps from a particular dataset may be displayed in a client's WWW browser.
- Queries and simple analyses (including map algebra, overlays, and vector-to-raster conversion) may be performed prior to downloading data by running a GIS as a CGI program.
- Downloading of data layers in the Spatial Data Transfer Standard (SDTS) format.

For example, if presented an image map of the United States, suppose that the client selects Texas. This leads to a county map of Texas. At this point, the client may select either a database for the entire state or for a particular county. Let's say that the client selected Upshur County, Texas and that no smaller region is defined on the GIS data server. The client would then be given a list of maps available for this county. For the sake of this example, assume that a vector map of streams and a raster map of land use are available. The client could then start a GUI for a GIS (as a CGI program) and then display these maps or perform a simple analyses:

- Convert the vector map to a binary raster map.
- Define a new raster layer using map algebra, perhaps to check a USGS streams map against a Thematic Mapper land use classification:\\ streams.tm=if (streams.usgs,landuse.tm)
- Display the new map.

After browsing data in this manner, suppose that the client deems the streams map valuable enough to make a personal copy. This vector map could then be downloaded from the server in SDTS format.

This type of interaction via WWW provides an excellent interface as well as world-wide access to a GIS data server. The elements of such a server necessary for the above example were developed. Implementation of each element is described below. First, a CGI interface to GRASS Lite, a graphical user interface to the GRASS GIS, as well as a display-only map-making interface, are described. Next, examples of data presentation/organization using image maps are given. Finally, a SDTS data conversion/transfer system is presented.

All of these interfaces are available at the following URL: http://ingis.acn.purdue.edu/

Integration of GRASS and the WWW

GRASS is primarily a raster-based system, originally developed by the U. S. Army-Construction Engineering Research Laboratory (USA-CERL). GRASS data sets, comprised of site, vector, and raster data, are arranged by location. Each location can have many mapsets or workspaces, perhaps arranged by themes or file ownership. The PERMANENT mapset in GRASS contains the system files defining the geographical extent and the projection information for the location.

The end-user typically does not have permission to modify the files in this mapset.

GRASS was chosen for this project because of many factors: (1) open file formats (Gardels, 1993 and Ireland, 1995), (2) freely-available source code (which allows for modifications and/or customizations), and (3) lack of licensing restrictions. As a public domain system, a WWW server using GRASS does not utilize costly floating licenses. In addition, one of the aims of our GIS/WWW project was to redistribute data developed for other research, which existed in GRASS format.

As with similar software systems, graphical user interfaces (GUIs) have been recently developed to allow the user to perform the GRASS commands in a graphical and user-friendly environment. GRASS Lite (Zhuang and Engel, 1995) is a GRASS GUI developed by Xin Zhuang of Wyle Laboratories (Arlington, Virginia) using the Tcl/Tk toolkit (Ousterhout, 1994). Tcl/Tk is native to the X Window System, `a vendor-neutral, system-architecture neutral, network-transparent windowing and user interface standard (MIT, 1991).' Since the X Window System is a network-transparent system, graphical applications can be physically run on the CPU of one machine but displayed on another machine's monitor (perhaps located on another continent), as long as both machines are on the Internet and are running the X Window System.

Using the Common Gateway Interface, software was developed to integrate GRASS Lite on the Internet. To utilize GRASS Lite for display/analysis of data through the Web, clients page through a series of three documents, exchanging information with the server. Each document is created dynamically by a CGI program. The first document uses the FORMS option in the HTML language to allow the user to specify an X Window System display for graphical output. It also instructs the user to allow the server access to their display (using xhost).

The next document allows the user to select a GRASS data set (LOCATION). It also overcomes a restriction that GRASS not be run concurrently by individual users (a built-in safeguard of GRASS for database integrity). To allow multiple sessions, a temporary HOME directory is then created with a GRASS start-up file (.grassrc) and a data directory. A data directory has a symbolic link to the PERMANENT mapset and a sub-directory containing a default mapset named workspace. Clients have write-access to their workspace but read-only access to the PERMANENT mapset.

The final document presented to the user executes GRASS Lite and provides helpful information to begin using the GUI as well as an e-mail address of the maintainer. As clients progress from the first to the third document/CGI program, information obtained from prior FORMS options are passed as hidden options in the HTML document. Each gateway program looks for specific items and encodes them in dynamically-created HTML documents.

In addition to map display, this GIS GUI on WWW allows clients to query GRASS databases and perform algebraic manipulations of raster data before downloading. Additional functionality may be provided in future versions. Figure 1 shows an example GRASS Lite session that was initiated via the WWW.

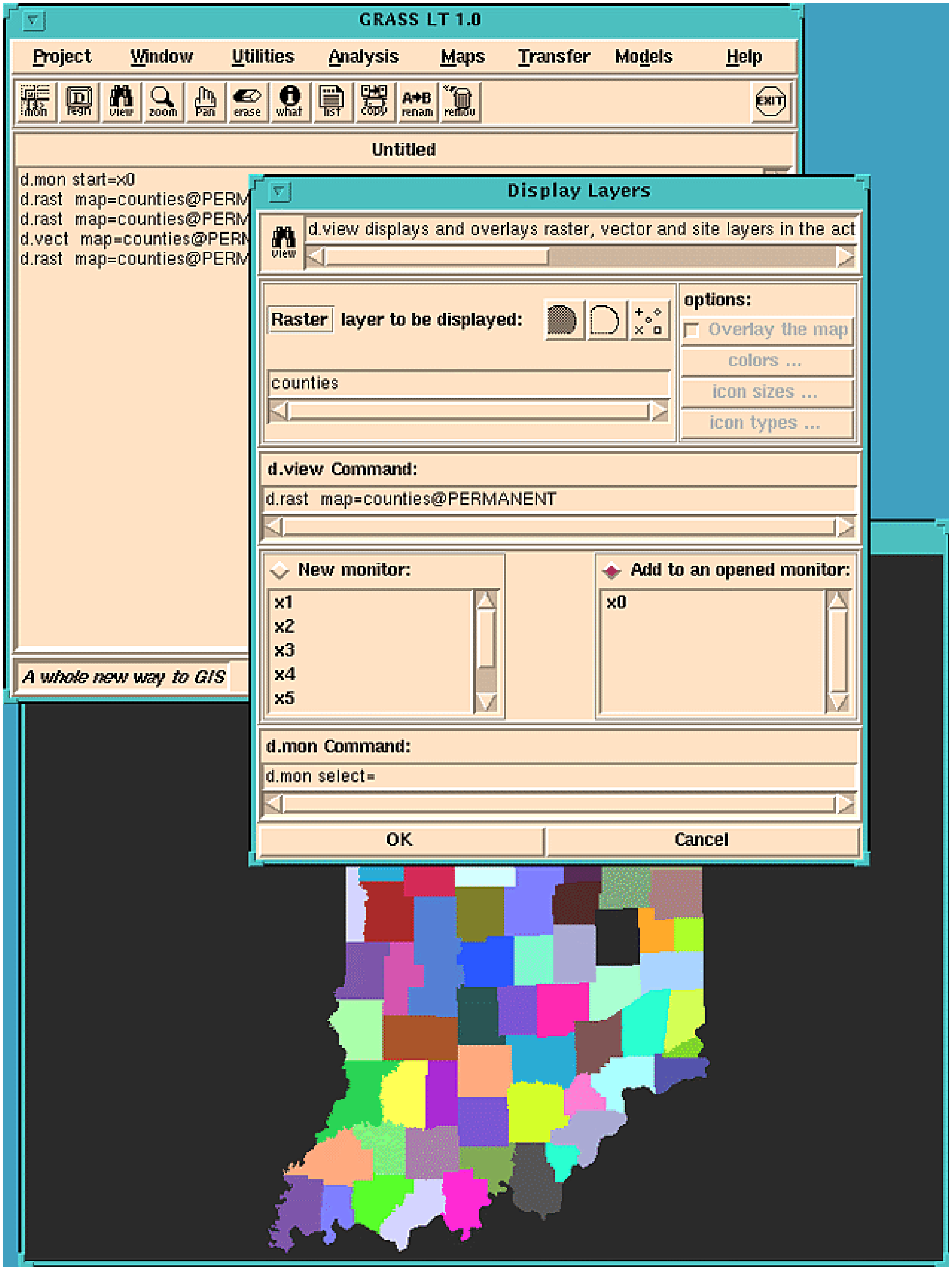


Figure 1. GRASS Lite Session via a WWW Session

It should be pointed out that the clients actually access a version of the software that was modified such that obvious security holes have been removed (e.g., any GRASS Lite options that gave users access to the UNIX shell have been disabled). More complete versions of this software are available from the author, but these are not recommended for use via WWW without security-related modifications.

The previous example using GRASS Lite requires users to run the X Window System, which is not always readily available to most PC users connected to the Internet. If users only need to view data (and not perform any types of geographic analyses), a display-only system (platform-independent) would be useful. Because of this, a similar approach was followed (using CGI scripts and the GRASS software) to build a display-only system. This system is also accessible from the above URL.

After selecting the GRASS data set, the LOCATION is posted to a CGI script that reads the available raster, vector and site data layers from an ASCII file and creates a HTML document with forms. The user has the option to select one raster layer, multiple vector layers and one site layer. The selected data layers along with the location name are posted to another gateway script which creates a HTML document that allows the user to select options to compose the final map. The options specified are the functionalities available in ps.map (Carlson, 1994), the PostScript cartographic output program of GRASS. The selected data layers and the corresponding map compositions are sent to another URL which processes the arguments and develops a script file that can be redirected into ps.map. The script reads the PostScript file created and converts it to a format viewable by most Web browsers. The final raster image is displayed in a HTML document as shown in Figure 2.

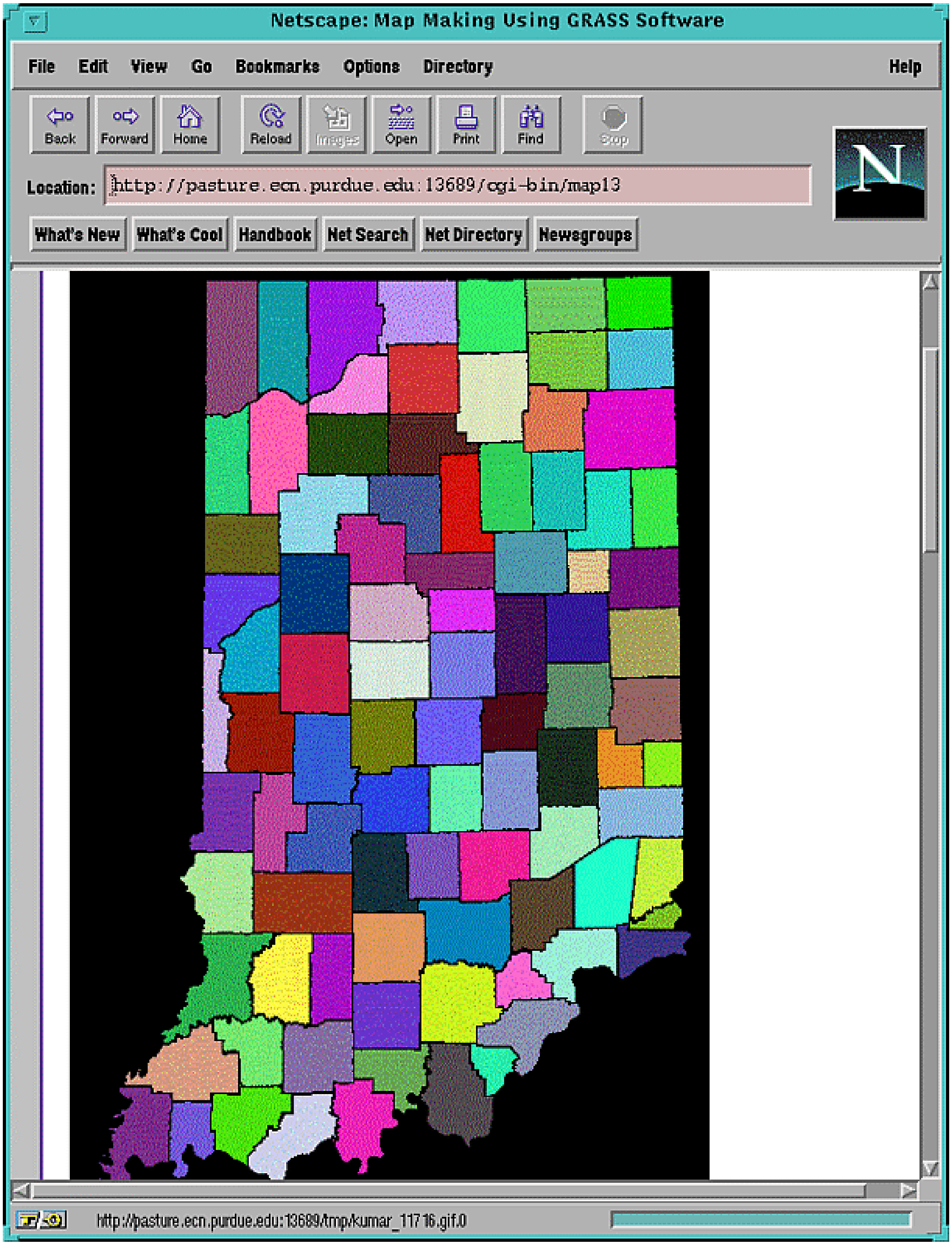


Figure 2. Result of Accessing and Displaying the Indiana Data Set County Raster Map

The approach described above can take a significant length of time (more than two minutes) to generate the desired map information and have it returned to the client Web browser (return time depends on the speed of the Internet connection). To overcome this constraint and to reduce the computational load on the server, the GRASS commands that display information were modified to directly create a file in gif format. The gif file can be directly displayed in the client WWW browser. A WWW form-based interface was written to provide access to this revised version of the GRASS display commands. The user selects the map or maps to be displayed, selects colors to be used, and provides text to be used as titles. This approach reduced the server computation time for creating requested maps by an order of magnitude as compared with the above approach.

Examples of GIS Documentation on the Internet

One of the interesting characteristics of the Web is that information is dynamic. This characteristic can be used to the advantage of programmers when software documentation is made available via the Web. Software documentation can be updated immediately after finding an error and users do not have to wait for the next version of the software for corrections. Users also do not have to use their own disk to store on-line manuals, which can be a substantial savings depending upon the amount of documentation available.

For the GRASS GIS (as well as most UNIX-based software systems), documentation is usually provided in the form of man pages in roff format. This format is versatile in that it may be read on both ASCII and graphics devices. In addition to being the native markup language of Web browsers, HTML also has this characteristic. As demonstrations of the utility of HTML with respect to man pages, many have written conversion utilities for roff-to-HTML. GRASS man pages, however, were written for a slightly different set of roff macros than those used by most UNIX software. Therefore, a custom conversion program was written for GRASS man pages.

The result was WWW-based documentation for GRASS commands. In addition to the benefits described earlier, WWW-based documentation ultimately becomes indexed in large search systems (e.g., WebCrawler by America Online and Lycos (http://www.lycos.com/). This indexing points potential users to GRASS when searching WWW pages (e.g., a search for map algebra may point clients to GRASS).

Examples of Data Presentation and Organization

Image maps were used to organize a visual search system for spatial data. Political and watershed boundaries of available data sets were drawn in a raster image. Selecting a coordinate in this region-sensitive graphic takes the user to either the corresponding data set or a metadata page describing the data. Below this image map is a link to data covering the entire region (e.g., the whole state).

An example illustrating this type of visual search mechanism utilized TIGER data for the State of Indiana. Since TIGER data are stored by counties, an image map of county boundaries in Indiana was created. By selecting the county of interest, a client could download all TIGER line data for that county.

Examples of Data Conversion and Transfer

On July 29, 1992, the Spatial Data Transfer Standard (SDTS) was approved as a U.S. Federal Information Processing Standard (NIST, 1992). As of February 15, 1993, U.S. governmental agencies are required to use this standard data format. Because of this requirement, most GIS packages, including GRASS, can now read and write data in this standard format.

The SDTS format was chosen in the development of a mechanism to make GRASS data available to users (who may or may not have GRASS, but whose chosen GIS will likely be able to import data in this format) through WWW. A series of CGI programs, similar to those described in the previous sections, allow clients to select any vector map in any available GRASS data set. The final CGI document runs the GRASS command v.sdts.out (Stigberg and Qian, 1995), creates a tar archive of the results, and sends the tar file back to the client. The first line of the output for this is:

```
Content-type: application/x-tar
```

Web browsers configured to recognize this type of MIME-encoding will present users with a dialog box asking for a file name to save the data.

Security Issues

One issue of concern for interfaces like these presented is system security. Some interfaces might become a security hole in the server system. Using temporary directories and/or a unique login (with restricted access) to run CGI scripts limits potential problems. Stand-alone systems with restricted access or password authorization might also reduce the risk of security problems.

As with other data handling procedures on the Internet, such as FTP, gopher, etc., it is always advisable for users to be aware of the data they are downloading, and more importantly, that others may be able to peek at data streams. For users of these interfaces, this may be of little concern since geographic information downloaded from public servers is rarely sensitive in nature. However, for sensitive data, encryption procedures (such as PGP (Zimmermann, 1995)) are an option.

SUMMARY & CONCLUSIONS

Several World-Wide Web (WWW)-based interfaces for GIS data sharing and software access were discussed. The GRASS Lite graphical interface to the GRASS GIS was made available via WWW using the CGI interface. This facilitated access to the geographic data sets and allowed simple analyses to be performed without actually downloading data or software. A platform-independent, display-only map creation interface provided a good browsing facility for potential data consumers. The data conversion capabilities of GIS were used to demonstrate the possibility of sharing the data in different formats, particularly in the Spatial Data Transfer Standard (SDTS) format. Presentation of spatial data using image maps allowed users to reach particular data sets quickly and efficiently. These interfaces demonstrated the capacity to view, manipulate, and distribute geographic data via WWW in an efficient, organized, and user-friendly manner. These interfaces are a major advancement in information sharing for GIS data.

FUTURE WORK

Future work may evaluate additional GIS-related applications via WWW. Automated techniques for locating, accessing and using disparate spatial data (or information derived from these data) that are distributed on the Internet are needed. Such techniques will facilitate the development of complex decision support systems and GIS applications that are capable of providing the information required to assist in solving a wide variety of problems. In the more immediate future, other similar interfaces can be developed to extract data from RDBM systems. Other GIS software, such as ARC/INFO, can be integrated on the Internet and the usage of this facility can be restricted to a particular group by taking advantage of the access authorization facility in HTTP. Access authorization and payment mechanisms may also be used to restrict access to particular data to support costly data development, software licenses, and system maintenance.

LIST OF ACRONYMS

ASCII	American Standards Code for Information Interchange (ISO 646)
CGI	Common Gateway Interface
FTP	File Transfer Protocol
GIS	Geographic Information System
GRASS	Geographic Resources Analysis Support System
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
HTTPD	Hypertext Transfer Protocol Daemon
MIME	Multipurpose Internet Mail Extension (RFC 1341)
PERL	Practical Extraction and Report Language
PGP	Pretty Good Privacy
SDTS	Spatial Data Transfer Standard (FIPS 173)
TCP/IP	Transmission Control Protocol/Internet Protocol
TIGER	Topographically Integrated Geographic Encoding and Referencing (U.S. Census Bureau)
URL	Uniform Resource Locator
WWW	World-Wide Web

REFERENCES

Albiltz, P. and Liu, C. 1992. DNS and BIND, A Nutshell Handbook, Mar 1993 edn, O'Reilly & Associates, Inc., Sebastopol, Calif.

Berners-Lee, T. 1994a. HTTP: A protocol for networked information, Internet Draft. Internet Engineering Task Force.
URL: http://www.w3.org/hypertext/WWW/Protocols/HTTP/HTTP2.html

Berners-Lee, T. 1994b. Uniform resource locators, a syntax for the expression of access information of objects on the network, Internet Draft. Internet Engineering Task Force. W3 Consortium and MIT Laboratory for Computer Science, 545 Technology Square Cambridge, Massachusetts.
URL: http://www.w3.org/hypertext/WWW/Addressing/URL/

Berners-Lee, T. and Connolly, D.W. 1995. Hypertext markup language - 2.0, Internet Draft. Internet Engineering Task Force. W3 Consortium and MIT Laboratory for Computer Science, 545 Technology Square Cambridge, Massachusetts.
URL: http://www.w3.org/hypertext/WWW/MarkUp/html-spec/

Borenstein, N.S. and Freed, N. 1992. Multipurpose internet mail extension (MIME), Internet RFC-1341, Internet Engineering Task Force.
URL: http://www.w3.org/hypertext/WWW/Protocols/rfc1341/0_Abstract.html

Carlson, P. 1994. ps.map: software for cartographic map creation. GRASS 4.1 Reference

Manual, U.S. Army Corps of Engineers, Construction Engineering Research Laboratories, Champaign, Ill.

Comer, D.E. 1995. The Internet Book, Prentice Hall, Englewood Cliffs, N.J.

Gardels, K. 1993. What is open GIS?, GRASSCLIPINGS: The Journal of Open Geographic Information Systems 7(1):40.

Hunt, C. 1992. TCP/IP Network Administration, A Nutshell Handbook, May 1994 edn, O'Reilly & Associates, Inc., Sebastopol, Calif.

Ireland, E. 1995. Data access is the path to mainstream mapping applications, Geo Info Systems 5(5):61-62. OpenGIS Special Section.

McCool, R. 1995. The common gateway interface, Software available from the National Center for Supercomputing Applications at the University of Illinois in Urbana-Champaign.
URL: http://hoohoo.ncsa.uiuc.edu/cgi/overview.html

MIT 1991. The X Window System, version 11, release 5 edn, Massachusetts Institute of Technology.

Montulli, L. 1995. Lynx users guide version 2.4, Software available from the University of Kansas.
URL: http://www.w3.org/hypertext/WWW/Lynx/Status.html

NCSA 1995a. NCSA httpd 1.4, Software available from the National Center for Supercomputing Applications at the University of Illinois in Urbana-Champaign.
URL: http://hoohoo.ncsa.uiuc.edu/docs/Overview.html

NCSA 1995b. NSCA mosaic, Software available from the National Center for Supercomputing Applications at the University of Illinois in Urbana-Champaign.
URL: http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/

Netscape Communications Corporation 1995. Welcome to Netscape, Software available from Netscape Communications Corporation, 501 E. Middlefield Rd., Mountain View, California.
URL: http://www.mcom.com/

NIST 1992. Spatial data transfer standard, Federal Information Processing Standard Publication 173. National Institute of Standards and Technology, U.S. Department of Commerce.

Ousterhout, J.K. 1994. Tcl and the Tk Toolkit, Addison-Wesley.

Postel, J. and Reynolds, J. 1985. File transfer protocol (FTP), Internet RFC-959, Internet Engineering Task Force.

Stigberg, D. and Qian, T. to appear 1995. v.sdts.out: software for exporting SDTS data. GRASS 4.2 Reference Manual, U.S. Army Corps of Engineers, Construction Engineering Research Laboratories, Champaign, Ill.

Zhuang, X. and Engel, B.A. 1995. Tcl/Tk GUI toolkit offers cross-platform application development, GIS World 8(7):58-60.

Zimmermann, P. 1995. The Official PGP User's Guide, MIT Press.

AUTHOR INFORMATION

James Darrell McCauley, Engineer
(formerly with Purdue University)
Case Corporation
email: mccauley@mcs.com
Telephone: 708.887.2055

Kumar C. S. Navulur, Research Associate,
Agricultural and Biological Engineering
Purdue University
W. Lafayette, IN 47907-1146
email: navulur@ecn.purdue.edu
Telephone: 317.494.1196
Fax: 317.496.1115

Bernard A. Engel, Associate Professor,
Agricultural and Biological Engineering
Purdue University
West Lafayette, IN 47907-1146
email: engelb@ecn.purdue.edu
Telephone: 317.494.1198
Fax: 317.496.1115

Raghavan Srinivasan, Agricultural Engineer and Associate Research Scientist
Blackland Research Center
Texas Agricultural Experiment Station
Temple, TX 76502
email: srin@brcsun0.tamu.edu
Telephone: 817.770.6670
Fax: 817.770.6678

A Network-Accessible Repository for the Characterization Of Spatial Ecosystem Components

Michael Sweet, Ray Ford, Ron Righter

ABSTRACT

Recent trends in network access, access tools, and user expectation are tempting public agencies and research labs to reorganize their information holdings to promote better internal and external access to datasets, dataset descriptions, and data transformations. A number of research projects are developing techniques to promote the construction and use of information repositories, but it is difficult to define the "ideal" characteristics of an information repository and build a system that satisfies a widely disparate set of users. Information environments typify complex systems in that they have physical, social, and mechanical components. Supporting an on-line data repository for disparate types of information involves interaction among all three of these components. Our work uses object-oriented techniques to address these interactions in the design of a repository aimed at sharing spatial datasets, non-spatial data, and modeling components in ecosystem research. Our implementation efforts are now sufficiently advanced to permit a certain amount of self-evaluation about the successes and limitations of our efforts. The object-oriented modeling paradigm is an elegant way to identify and classify the essential attributes of complex physical systems in our target domain. However, it has been less easy to identify and implement the functionality required for a repository to be "distributed and open, yet reliable and secure," given rapid changes in networks and network access tools.

INTRODUCTION

Model-building is an interdisciplinary, integrative learning activity that must operate within the constraints of a network systems problem and a synthesis systems problem (Patten 1994). The network systems problem is the ability to elucidate the "unexpected, unpredictable, and inexplicable consequences of cause whose origins are lost in the complexities of labyrinthine webs." The synthesis systems problem focuses on "the ability to meld together multi-disciplinary knowledge and perspectives on complex problems." For ecosystem research, the task in resolving the network systems problem is one of developing appropriate investigative and modeling methodologies. Resolving the synthesis systems problem means promoting physical, mechanical and social structures that allow a team of specialists with varying backgrounds to realize an interdisciplinary, collaborative approach to ecosystem research.

The designer of an information system intending to meet the challenges of ecosystem modeling needs to be cognizant of the role an information system plays in the synthesis problem described above. Information system designers must be equally aware of the complexity that physical, social and mechanical systems impose on the activity of modeling, and how changing technology can affect the boundaries of these systems. The Ecosystem Information System (EIS) design intends to encourage users from a wide range of organizations to share metadata descriptions and register dataset holdings in an extensible, network-accessible repository. EIS is an information system that attempts to bridge some of the social, mechanical and physical boundaries alluded to by Patten. We report on the key technical aspects of EIS design, as well as its implementation status and still evolving concerns.

DISCUSSION

Recent trends in network access, access tools, and user expectation have led public agencies and research labs to reorganize their information holdings to promote better internal and external access to datasets, dataset descriptions, and data transformations. The presumed goal is better support for regional planning and policy formulation. This extends information access beyond modelers and scientists to include policy makers, managers, agency officials, organizations and the public at large. Developers need to consider this type of broad-scope information access in designing information systems and repositories. By its very nature, broad-scope access must be based on decentralized, vendor independent, widely available hardware and software.

The body of knowledge in any modeling enterprise is dynamic, depicting an ever changing collection of both new and established representations. Interdisciplinary modeling efforts, such as those that typify ecosystem modeling, tend to be particularly complex. For example, Davis (1995) provides a user's perspective on the salient requirements for an information system that would support a biodiversity database containing both spatial and non-spatial elements. Davis observes that "compiling such disparate data even for a single location can be frustrating because of the varying disciplinary conventions for data production, documentation and exchange." Many conflicting schemes for data representation by software vendors further complicates this task. Recognizing and planning for the diverse and transient aspects of this knowledge-base is critical to designing the "data representation" component of a supporting information system.

The physical, mechanical and social systems that surround modeling efforts present additional obstacles to the development of an effective information repository. The manifestations of these compelling influences appear at a coarse design level as factors in the human-computer interface, and at a fine level in the representation of data elements. For example, the geographic separation of scientists working on a project, which can range from meters to hundreds of kilometers, is an important aspect of the social framework of ecosystem research. The behavior and expectations of scientists, the cognitive aspects of human-computer interaction, and the capacity for scientists to map their problem domain to the abstract constraints of digital environments further define the social system. The designer of a robust data repository must fully consider the influence of these social, mechanical, and physical systems when making design choices. This is particularly true as computer networks and distributed information systems evolve as mechanisms to further facilitate social systems in bridging geographic and organizational boundaries.

EIS Overview

Rather than collecting and maintaining a collection of data in a centralized location, an EIS repository provides a highly structured index into a collection of information of various types that is assumed to be distributed among a set of contributors. The sociological distinction is important. EIS grows not through the collection and control of information resources, but by publicizing the availability of information from contributors. That is, EIS allows contributors to register the types of information they are willing to share and build flexible index structures to best organize their information, without asking contributors to relinquish physical control over the information. The rationale for this indexing approach stems from the realities of cross vendor/platform support, data volume, data ownership and distribution concerns. There is no environment that fully encapsulates critical details in data or program description (Pancake 1995), and there is no repository that could possibly hold all the available information. More importantly, owners perceive datasets as information capital, and even if these assets are shared the owner may not wish to relinquish all control.

This new type of data repository has become feasible only recently with the emergence of widely available, standardized computer networks and network access tools. What seemed a radical concept only a few years ago is now widely recognized as the "web-work of publicly accessible data" paradigm embodied in the World Wide Web. The tremendous growth in popularity of the Web in a very short time typifies some of the complexities the repository designer must resolve. The success of the Web demonstrates the feasibility of broad-scale information sharing and dramatically raises public expectations. In fact, the Web is so dominant that the public expects that any information sharing environment will BE the Web, not just BE LIKE the Web.

WEB Interface

The EIS software design has always embodied a 'web-like' sharing paradigm, with the implementation utilizing existing and familiar network information access tools. However, as scientists and the general public become 'net-aware', the Web dominates their expectations about information access and the information system interface. The growth of popularity in the Web has taken place concurrently with the design and implementation of EIS, forcing the EIS designers to adapt their goals to meet emerging public expectations. The result is that EIS now has two primary interfaces. For simple browsing and searching, users can access EIS information through standard Web/WAIS tools (Fig. 1). Thus, Web users gain full browse and search access to EIS contents, without having to install or even know much about EIS. However, people who want to contribute to the distributed EIS database need to install EIS and use the native EIS interface to gain the full range of services. This condition is similar to the realities of the Web -- those who install and maintain Web servers need to know considerably more about the Web as a software system than those who simply browse for existing information.

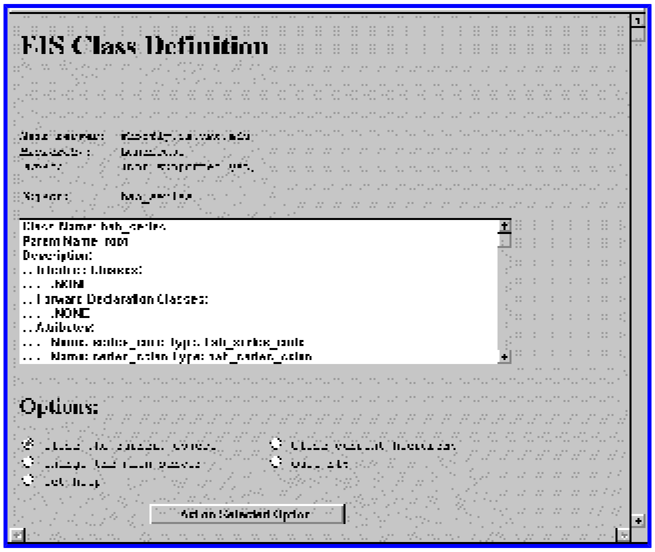


Figure 1. EIS Web interface.

The native EIS interface is in essence a "mechanical" requirement. A contributor to EIS must use this interface to guarantee that certain information is provided in a coherent manner. In contrast, the Web interface to EIS is a social enhancement; it broadens access to the EIS repository and invites the participation of a clientele that spans a wide range of technological, geographic, and organizational boundaries.

EIS Data Model

The architecture of the EIS repository starts with a distinction among three key concepts: data model, data value and type, and data structure (Righter 1994). In general, a *data model* is an abstract or symbolic representation of reality defined in terms of a set of specific properties or *attributes*. The domain experts in ecosystem modeling define the abstractions and properties that best represent their domain knowledge. Data values, drawn from a *data type*, represent the specific values of the attributes of a data model. Data types define a template for a legal set of data values. Domain knowledge from the target domain or 'outside' domains contribute the data types. For example, geographers may best describe the appropriate data type for a given spatial phenomenon in ecosystem modeling. *Data structures* are specific techniques to organize and represent a collection of data values. In an interdisciplinary research environment, domain experts may draw on the experience of a computer scientist to define the appropriate data structure to represent a specific data model.

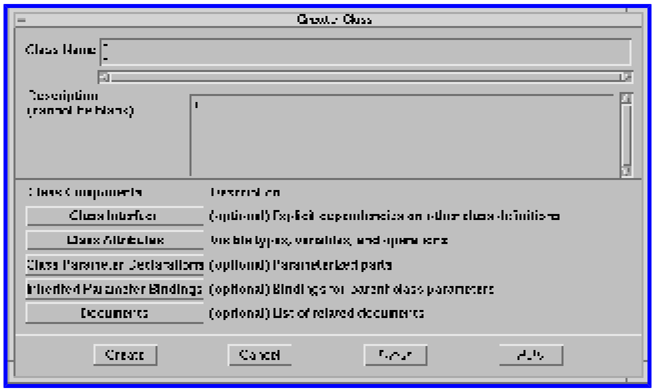


Figure 2. EIS interface for data model description form.

It is this collection of data attributes, values and types as symbolic abstractions of reality that represent domain knowledge and serve as fundamental modeling components (Fig. 2). Within ecosystem applications, these abstractions typically represent spatial phenomena. For example, consider the digital elevation data model in USGS DEM format. The DEM data model locates one elevation value at each point in a regular grid. The DEM data structure describes the logical mapping of a list of actual elevation values to the data model (Righter, 1994). What is missing from this representation is a mechanism to describe the relationship between abstractions. Papazoglou (1995) refers to this as the 'conceptual schema' or the 'intension of data'. Papazoglou discusses the distinct features of intensional and extensional schema. We gain knowledge in an extensional framework by evaluating or querying data elements. In an intensional framework, we gain knowledge from how we represent and structure data elements as a collection of abstractions.

Since EIS is an index to data, EIS is primarily an intensional framework. The goal is to structure knowledge or metadata rather than just structure information at the level of a data element. The metadata based indices thus provide a well-ordered intensional framework, suggesting that users will find this an efficient mechanism for accessing the appropriate subset of data for extensional queries. Papazoglou claims that people who employ symbolic representations and intensional queries will ultimately produce more meaningful extensional queries. Papazoglou recommends the abstract data-typing facilities and inferencing abilities of the object-oriented data model as the appropriate model for an intensional framework.

By structuring the data repository around an object-oriented design the local repository builder can construct his/her own classification system, or apply or modify an existing classification system depending on his/her specific interests. The process of identifying and organizing relevant information for a repository begins with an extensive domain analysis (Righter 1993). The domain analysis seeks to identify a key set of concepts, entities and relationships between entities within an application domain like ecosystem modeling. If the domain analysis is robust, the characterization will apply to a wide-range of users and likely apply to other application domains. In an interdisciplinary environment, each domain expert can have a role in defining EIS objects that contribute to the overall intensional framework of the data repository. If we expect to attract users, EIS must contain elements of value to other ecosystem researchers. EIS must exemplify the process through which potential users can re-cast their own datasets and models into the object-oriented paradigm.

Class hierarchy

The EIS repository is structured as an object-oriented class hierarchy. Formally, a *class* within the hierarchy represents a type of data source or data model. Informally, a class description represents a metadata description that defines the set of logical characteristics, or attributes, possessed by each member of that class. A class attribute can describe either a data property or an operation on an instance of that class. An *instance* is simply a member of a class, i.e., an actual dataset consistent with the metadata characteristics defined by its parent class (Fig. 3).

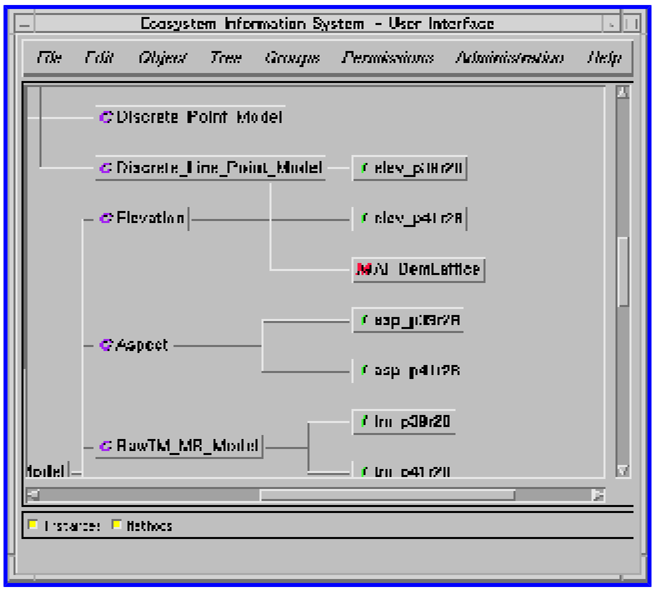


Figure 3. Example EIS class hierarchy.

We refine and extend the class hierarchy by describing a key relationship between classes. This relationship arises whenever we can define a new class from an existing class in such a way that the new class has all of the attributes of the existing class as well as one or more new attributes. This process of specialization is referred to as creating a new subclass from an existing parent class, such that the subclass inherits all the attributes of the parent. It is this process of successive decomposition that defines and extends the class hierarchy. Also, by explicitly recording the parent-class/instance relationship, we begin to unambiguously organize collections of dataset instances into a manageable hierarchy (Righter 1994). It is contextual meaning, rather than just content, that classifies and organizes objects. Papazoglou refers to this as the semantics of taxonomy. Note that generalization, in the form of a bottom-up approach to classification, is the inverse of specialization.

The principle of attribute inheritance is a simple, yet effective way to describe the similarities and differences of various data sources that evolve from a common ancestry. The class/subclass relationship forms a class hierarchy which embodies the intensional relationships. This class hierarchy should stimulate a deeper understanding of the application domain and answer the question of "What is it?" for a given instance.

Browsing an EIS class hierarchy is akin to conducting a visual query into the class/subclass relationship, always starting at the topmost (i.e., most general) class. Objects that are semantically similar will appear in close proximity within the hierarchy. Interpretations made as a result of the relationship of classes within the design space is indicative of knowledge gained from the intensional aspects of the class hierarchy. When a user views the class hierarchy through the Web interface, hierarchical structures appear in a textual, indented-outline form. Individual classes appear in their textual (syntactic) form, and individual datasets or programs appear with a textual location reference (Fig. 4).

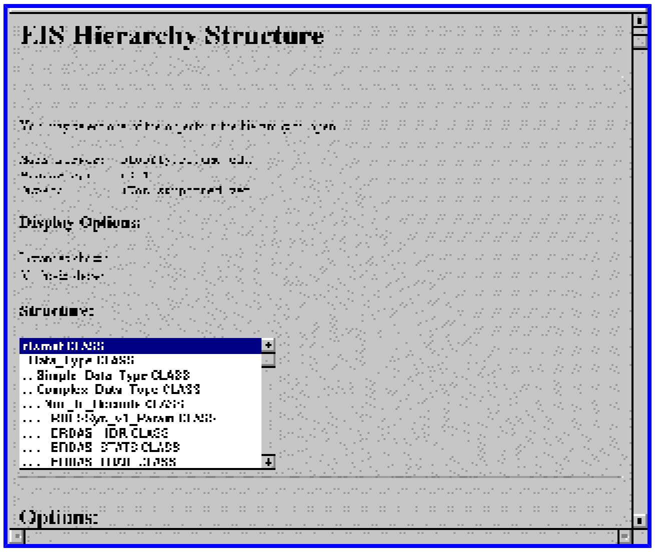


Figure 4. EIS Web interface to a class hierarchy showing classes, instances and methods.

The EIS repository database is a collection of class hierarchies. That is, a user can extend an existing hierarchy to classify his/her data, develop a new hierarchy, or both. Because the terminal nodes in a hierarchy are references to datasets, not the datasets themselves, there is no reason that a given collection of data cannot be indexed in more than one fashion. In fact, for most interesting collections of datasets we would expect multi-indexing to be the norm. Each indexing scheme would be represented by a separate class hierarchy.

SELF-EVALUATION

Implementing a vision of a distributed 'Ecosystem Information System' in a period of astronomical growth and awareness in the Internet and the proliferation of Internet-based tools is a challenging and sometimes daunting task (Ford 1994). Regardless of these rapid changes, the fundamental concepts and goals of EIS remain in place. There are still social as well as technological barriers to overcome, even if we assume that the EIS software is the appropriate tool for realizing those goals.

The biggest implementation barrier is to stabilize the EIS code and construct a core database that will attract the participation of other users. We view the Web interface as a critical component in promulgating EIS. The most obvious technological and social barriers looming are those involving customized, proprietary net browsers and searchers. To allow more complete functionality, EIS should properly base many enhancements on "Web extensions".

Participation in EIS from the "data owner" standpoint requires clarification of the EIS paradigm and demonstrating that EIS provides a way to structure information in a manner that makes sense to them. It is important to emphasize that EIS offers an accessible "public" way to extend and refine a user's classification system, while allowing the user to retain control of the data. We think data owners will participate fully in EIS once they recognize the value in a public classification of data combined with local control. Objected-oriented design methodologies are most efficient when specialist with similar interest share knowledge. Any social or technological framework that directly supports interaction in the design portion of the class hierarchy will strengthen the EIS repository.

While a robust and user-friendly implementation of the EIS software is important, the content of the initial database is critical to long-term success. The repository must present information of sufficient interest to the new user and illustrate the concepts and value of the object-oriented paradigm. The initial repository must serve as a model if users are to gain interest in re-casting their own work into this framework.

The intensional framework of the class hierarchy is critical to the user's ability to understand the organization of the data repository. Providing a concise methodology for structuring information will reduce search time, and results in more meaningful results from intensional queries (Papazoglou 1995). Many Internet-based repositories lack user involvement in the structuring of domain knowledge in a way that is meaningful and familiar to the user. Papazoglou defines a full suite of intensional query tools that can add to the user's comprehension of the repository. For the EIS design team, investigating Internet-based approaches to intensional query tools for indexing and searching will undoubtedly extend EIS from a framework for representation to one of understanding.

CONCLUSION

Patten's discussion of the network systems and synthesis systems problem emphasizes the importance of the intensional characteristics of the EIS repository. We can probably find many examples and possibly failed examples of efforts to build large collections of disparate data. As we move into the 21st Century, we need to examine previous efforts to organize large collections of disparate data and identify the benefits and uncertainties associated with different intensional and extensional approaches.

Patten also emphasizes the importance of collaborative, integrative modeling, even in the mundane task of populating a data repository. It is through shared experiences that we begin to develop a common working framework. Patten notes that this shared framework promotes "a willingness to (1) share knowledge with others, (2) absorb other's knowledge, and (3) incorporate acquired insights into the model under construction." In the laboratories and research projects in which we work, that trend is already apparent. EIS is a development framework to explore the practicality of addressing the complexities associated with a collaborative data repository. Evaluating EIS in the context of existing physical, social and mechanical systems can help set future direction in the design of shared data repositories.

ACKNOWLEDGMENTS

This work is partially supported by National Science Foundation grant BIR-9301760 and USDA Cooperative Research grant MONZ-9408.

LIST OF RELATED READINGS ON EIS

Ford, R., Righter, R., Duce, T., Hemige, V., and Thompson, D. (1995) EIS: A Network-Accessible Repository for Ecosystem Modelers and Managers. *Proceedings of 1995 ACM Symposium on Applied Computing, Nashville, Tennessee* pp. 76-80.

Reimer, Y., Ford, R., Wilde, N. (1996) Merging Task-Centered UI Design With Complex System Development: A Risky Business. *1996 ACM Computer Science Conference, Philadelphia, PA* (in press).

Ford, R., Hellenga, G., Palaiya, V., Thompson, D. (1996) Linking Specialized Network Data Repositories to Standard Access Tools. *1996 ACM Symposium on Experimental Computing and Applications Development, Philadelphia, PA* (in press).

Internet access to the Web interface to EIS.

REFERENCES

Davis, F.W. (1995) Information systems for conservation research, policy, and planning. *BioScience Supplement on Science & Biodiversity Policy* 45:1 pp. S36-S42.

Ford, R., and Running, S.W., and Nemani, R. (1994) A modular system for scaleable ecological modeling. *Computational Science and Engineering* 1:3 pp. 32-44.

Pancake, C.M. (1995) The promise and the cost of object technology: a five-year forecast. *Communications of the ACM* 38:10 pp. 33-49.

Papazoglou, M.P. (1995) Unraveling the semantics of conceptual schemas. *Communications of the ACM* 38:9 pp. 80-94.

Patten, B.C. (1994) Ecological systems engineering: toward integrated management of natural and human complexity in the ecosphere. *Ecological Modelling* 75/76 pp. 653-665.

Righter, R. (1993) An object-oriented domain analysis of ecosystem modeling. *M.S. Thesis,* *Department of Computer Science, University of Montana*.

Righter, R., and Ford, R. (1994) An object-oriented characterization of spatial ecosystem information. *Mathematical and Computer Modelling* 19:8 pp. 17-29.

AUTHOR INFORMATION

Michael Sweet
School of Forestry
University of Montana
Missoula, Montana 59812
Tel.(406)243-5265; Fax.(406)243 4845
Email: sweet@selway.umt.edu

Ray Ford
Department of Computer Science
University of Montana
Missoula, Montana 59812
Email: ford@cs.umt.edu

Ron Righter
Wildlife Spatial Analysis Laboratory
University of Montana
Missoula, Montana 59812
Email: rlr@selway.umt.edu

Paul Gessler, Neil McKenzie and Michael Hutchinson

Progress in Soil-landscape Modelling and Spatial Prediction of Soil Attributes for Environmental Models

Abstract

This paper summarizes progress and new work for the quantitative spatial prediction of soil attributes. An example spatial prediction of the soil profile carbon pool is presented for a study area in Australia. New methods integrating soil layer models and a hillslope profile sampling procedure are demonstrated for development of spatially-averaged hillslope models for three study areas. Visualizations of convergent and divergent hillslope soil layer cross-sections for each study area are presented with interpretations of landscape function.

1. Introduction

In Gessler et al. (1995) we reported on the development of generic soil-landscape modelling methods to provide explicit and quantitative spatial predictions of soil attributes that may be used in other, more comprehensive, environmental models. These methods build on the work of Moore et al. (1993), McKenzie and Austin (1994), McSweeney et al. (1994) and others. This paper provides: (i) a brief summary of the methods; (ii) discussion of progress and experiences in applying the methods over the last two years; (iii) an implemented example; (iv) a presentation of new work integrating the methods for development of hillslope models; and (v) a critical evaluation of the overall approach and requirements for additional components.

1.1 Summary of Methods

The approach outlined in Gessler et al. (1995) assumed that local terrain modifies climate and parent materials and therefore served as a simple surrogate integrating many landscape processes that influence soil patterns. Spatial analysis of a digital terrain attribute, the compound topographic index (cti), was used to define the local catenary/landscape population and catenary environmental gradients for allocating field sample locations. The assumption was that the stratifying variable, cti, captured the range and relative distribution of variation in the landscape for the attributes of interest. This was an attempt to explicitly quantify the implicit process often used in traditional soil survey. The strategy could be altered and adapted in many ways depending on the requirements of a particular project and availability of prior information.

After collection of field data, exploratory data analysis (Cleveland 1993) was used to search for correlations between the soil attribute being modelled and environmental attributes that were simpler to measure and available in a spatially continuous manner. Useful relationships were then confirmed and defined by statistical models. The statistical models were evaluated and improved by: (i) analysis of residuals; (ii) visualization of a spatial implementation of the model; and (iii) field verification. Generalized linear models (McCullagh and Nelder 1989) were used for spatial prediction of basic soil layer patterns (A horizon depth, E horizon presence/absence, Solum depth).

The first objective of the modelling was to develop empirically-based spatial predictions. The second objective was to interpret useful environmental correlations discovered in the modelling process for elucidation of underlying landscape processes responsible for the patterns.

1.2 Progress and Experiences in Application

The first study area (Griggward) was expanded and two additional study areas (Ladysmith, Brucedale) were sampled in the same manner described in Gessler et al. (1995). A broad range of soil attributes (physical, chemical and morphological) were modelled using a common set of procedures. Additional environmental variables (e.g. climatic, geochemical, digital imagery) were acquired over the study areas and sampled for use in predictive modelling. An example is presented below and comprehensive reporting of the study findings will be presented elsewhere.

The process used to develop statistical models described in Gessler et al. (1995) was improved using tools available in the S-PLUS (Statistical Sciences 1993) statistical computing package. For each individual soil attribute, plots of the univariate sample distribution and bi-variate relationships with sample depth and soil layer or horizon factor were produced. These indicated whether: (i) data transformation for normalization was warranted; (ii) if the soil attribute varied systematically or smoothly with depth in the soil profile; or (iii) if morphologically described soil layers strongly partitioned the variation suggesting modelling by individual soil layers. Conditioning plots or co-plots (Cleveland 1993) were routinely used to search for conditional relationships in terrain attribute space. Regression tree models (Breiman et al. 1984, Clark and Pregibon 1992) were also used to explore for possible conditional or complex nonlinear relationships in the data (landscape) that may be missed by other data exploration and model fitting criteria. A comprehensive stepwise exploration of both linear and nonlinear fits of all potential explanatory variables was conducted using the Akaike Information Criterion (Akaike 1974). This provided an objective numerical assessment that balanced the reduction in residual deviance with complexity of the model as defined by the degrees of freedom consumed (e.g. parameters fitted). Generalized additive models (Hastie and Tibshirani 1990) with scatterplot smoothers were routinely used for nonlinear fitting. If model residual patterns showed non-normal characteristics the "quasi" family was used in generalized linear or generalized additive models to allow use of stabilizing link and variance functions and relaxation of assumptions about the data distribution (McCullagh and Nelder 1989, Hastie and Tibshirani 1990).

In the end, our preference was to use simple models with linear terms if possible. There are several reasons for this. First, our longer term intention is to integrate the models for understanding pattern/process relationships within and between study areas. Secondly, subsequent work will investigate how we attempt to incorporate and account for estimations of error entered at various steps. Thirdly, we hypothesized that the further we push the models to fit the sample data by adding additional explanatory variables and higher order fits, the more local the model becomes and therefore less useful for prediction at regional scales. This will be tested in subsequent work.

The concept of a soil "type" was disregarded in this research. Instead, we focussed on the collection of useful datasets that can be continually analyzed as methods and our understanding develop. We explicitly stated the measurement methods, sample size and scale of each soil attribute and environmental variable (meta-data) realizing that they are taken over a range of different scales. Our spatial predictions assume the quantified relationships can be applied in the broader spatial context defined by explicit physiographic domains. Classification and imposition of traditional soil types or taxonomic classifications are not precluded in this approach, but our research focusses on the development of quantitative and continuous predictions of individual attributes with defined confidence and uncertainty based on collected sample evidence.

This approach enabled the development of a broad range of different models for individual soil attributes using a broad range of explanatory environmental variables and statistical modelling tools. Regression tree models often provided the largest reductions in residual deviance, but often produce stepped (e.g. non-continuous) prediction surfaces unsupported by field investigation. Soil layers proved very useful in partitioning the variation for many soil attributes, and hence supported the concept of a building block approach where soil layers were used to aggregate or disaggregate soil attribute data for modelling and spatial prediction where appropriate. Reductions in residual deviance ranged from 10 (A horizon exchangeable sodium percentage) to 90 (solum depth) percent, indicating the span of prediction certainty for different attributes that should be an explicit part of model implementation.

1.3 Example Integration of Models for Spatial Prediction

Total carbon was one of the soil chemical attributes measured in this study. In soils where negligible calcium carbonate is present, as was the case here, total carbon provides a measure of organic carbon held in the soil relating to biomass (e.g. microbial, macrobial, humus, plant material). The amount of carbon in the soil pool is an important component of the carbon cycle and knowledge of it is useful for broader biosphere modelling.

A plot of the bivariate relationship between total carbon and sample depth showed a very smooth relationship where percentage total carbon declined rapidly with soil depth. Conditioning plots using the compound topographic index as the conditioning variable showed that this relationship was invariant with landscape position. This indicated that a generalized additive model with a scatterplot smoothing spline could be used to model carbon over the landscape as a function of soil depth.

By combining models previously developed for solum depth and A horizon depth (incorporating E or A2 horizons) with the spline profile total carbon model and assumptions about soil bulk density, we developed a spatial prediction of the soil profile total carbon pool across the landscape. Bulk density was not measured as part of this work, but a recent survey (Geeves et al. 1995) encompassing the study area provided a regional mean A horizon bulk density (rhob) of 1.5 Mg m-3 of soil and a regional mean B horizon bulk density (rhob) of 1.6 Mg m-3. Using these, a predictive equation was constructed as follows:

Soil Profile Total Carbon =

$$\int_0^{adep} f(Totc.gam(s(dep))d(x) * \rho_b(Ahor) + \int_{adep}^{soldep} f(Totc.gam(s(dep))d(x) * \rho_b(Bhor) \qquad (1)$$

The integrals of total carbon for the A and B (B equals solum depth minus A horizon depth) horizons as predicted by the GAM model, computed on a one centimeter depth increment, is multiplied by the bulk density for the A and B horizons. Assuming that this model was valid across the area encompassed by a 400m2 grid cell, a spatial prediction was computed. Figure 1 shows a color coded spatial implementation of the modelled variation of the profile total carbon pool for the Griggward study area. A histogram equalized stretch was applied to enhance visual display, therefore the color legend does not have equal intervals of tons/hectare carbon. The total carbon GAM model provides a percentage reduction in residual deviance of 84%. The A horizon depth and Solum depth GLM models provide 78% and 77% reductions in residual deviance. This indicates that each of these component models can be predicted with a high level of certainty.

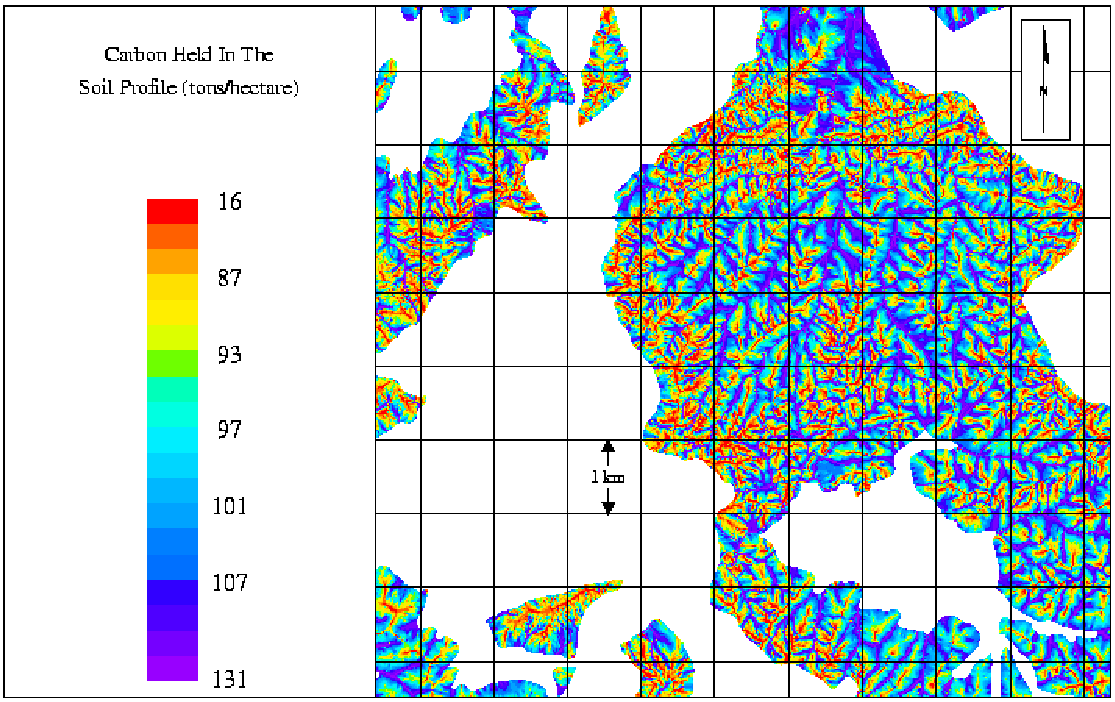


Figure 1. Predicted Soil Profile Carbon Pool (Griggward Study Area)

1.4 Hillslope Models of Soil-landscape Structure

The work presented thus far has focussed on the development of quantitative models for predicting soil-landscape patterns. The intention is to provide a good base from which to build an understanding of landscape processes to facilitate simulation of landscape dynamics. Forman and Godron (1986) and Turner and Gardner (1991) state that three important landscape characteristics are structure, function and change. Our work follows these concepts by first developing quantitative models describing the soil-landscape structure as defined by soil layer patterns. The soil layers, as discussed in McSweeney et al. (1994) and Gessler et al. (1995), are a result of pedo-geomorphic and hydrological processes and therefore can be used as a link or pointer to material and energy flow processes and pathways in landscapes.

In undulating landscapes with open drainage, the catchment or watershed is an important spatial area for understanding flow connectivity and flow accumulation that can be approximated using digital terrain attributes (Moore et al. 1993). The watershed is comprised, at a more local level, of a continuum of converging and diverging hillslopes (Carson and Kirkby 1972). In the past, hillslope block diagrams displaying soil layers have been a useful technique for visualizing hillslope patterns and flow pathways (Gerrard 1981, Walker and Butler 1983, Buol et al. 1989). However, these visualizations are often idealized two-dimensional hillslopes and not based on quantitative models developed over the three-dimensional soil-landscape continuum.

In reality, there is a complex mixture of landform curvatures causing flow to converge and diverge. Conceptually, the two end-members of this continuum would be hillslopes that continuously converge from summit to base and those that continuously diverge from summit to base. Everything else falls in-between as a mixture of converging and diverging hillslope elements. Therefore developing visualizations of these two end-members using quantitative models will be useful for interpreting landscape function and comparison of hillslope patterns between different study areas. The following sections describe methods and preliminary results for visualizing hillslope patterns for interpretation of material and energy flow processes in landscapes.

2. Material & Methods

2.1 Study Area Environmental Characterization

Three study areas were selected to cross a regional climatic gradient on two different parent materials. Detailed environmental characterizations of each area in the form of climatic, geologic, geomorphometric and airborne gamma radiometric spatial characterizations (e.g. summary statistics) have been created for each area and Table 1 presents six study area means indicative of basic physiographic differences. The climatic variables were developed using the ANUCLIM software (McMahon et al. 1995) output to a grid node spacing of approximately 245m. The geomorphometric variables were developed using the methods described in Moore et al. (1993) and Gessler et al. (1995) computed from a 20m grid node spacing digital elevation model (DEM) derived from a 1:25 000 topographic map source (10m contours).

Table 1. Study Area Spatial Mean Environmental Characterizations

	Climate			Geomorphometry		
Study Area	mean annual temp. (°C)	total annual precip. (mm)	mean annual radiation (mJ/m^2/day)	mean elevation (meters)	mean slope (percent)	mean compound topographic index
Brucedale	15.53	509	17.78	235	3.48	9.35
Ladysmith	15.11	575	17.61	259	8.68	7.75
Griggward	13.98	709	17.18	383	11.42	7.16

The Brucedale study area is 9 500 hectares in size (centered on: 147° 25'E, 35° 05'S) and situated on very gently undulating granitic parent materials covered by a thick mantle of aeolian clay. The dominant land uses are cereal cropping and pastoral grazing. The Ladysmith study area is 5 800 hectares in size (centered on: 147° 29'E, 35°13'S) situated on gently undulating hills of Ordovician metasediments dominated by slightly metamorphosed shales and sandstones of marine origin. The dominant land use is pastoral grazing. The Griggward study area (centered on: 147°27'E, 35°24'S) is 5 300 hectares in size situated on rolling hills of Ordovician metasediment parent materials. The dominant land use is pastoral grazing.

2.2 Soil layer modelling

Approximately 80 locations were sampled in each study area following the methods described in Gessler et al. (1995). A broad range of GLM, GAM and TREE soil layer models were developed using the procedures described above for A horizon depth, probability of E horizon presence/absence, E horizon depth and solum depth. These are the basic functional soil layers from which more detailed work may proceed. The Brucedale study area showed only two occurrences of E horizons, therefore E horizon models were not developed for this study area.

The best model for each soil layer attribute was selected based on reduction in residual deviance, degrees of freedom consumed and type of model. The general order of preference for model type was GLM, GAM and TREE. The selected soil layer models were implemented as spatial prediction surfaces using map algebra tools.

2.3 Hillslope Profile Sampling

The intention of the hillslope profile sampling was to develop a visualization of the average convergent and divergent hillslopes in each study area. This involved four steps of: (i) developing a quantitative definition of a hillslope; (ii) developing a sampling strategy to provide convergent and divergent hillslope datasets; (iii) editing and averaging the data in hillslope distance space; and (iv) developing display graphics.

A hillslope was defined as a spatial object that maintains flow connectivity from summit (hillslope initiation) to base (hillslope conclusion). Following empirical experimentation with digital terrain attributes, hillslope initiation cells were defined as those cells with less than two DEM grid nodes flowing into them. Hillslope conclusion cells were defined as cells with greater than 100 DEM grid nodes flowing into them. These usually corresponded with streamlines defined on digital 1:25 000 topographic map sheets for the area. Convergent and divergent hillslope components were then defined as the remaining cells with plan curvature greater than zero (convergent) and plan curvature less than zero (divergent). A convergent hillslope is therefore a hillslope that starts at an initiation cell and flows continuously through convergent cells downslope to a hillslope conclusion cell; and vice versa for a divergent hillslope. Using these principles, a visual display was created that showed the four-class grid overlaid with flow vector arrows and a 1:25 000 contour map (10m contour interval). Convergent and divergent hillslopes could then be traced to sample along the hillslope profile from each predicted soil layer surface.

Ten convergent and ten divergent hillslopes were sampled in each study area. A one by one kilometer line coverage was placed over each study area and each 1 km2 cell numbered sequentially. A random number vector was developed and sampled to randomly select ten 1 km2 subareas for hillslope sampling in each study area. This ensured no bias for any part of the three study areas. The data files generated were then edited to ensure that the hillslope distance intervals were identical for averaging. The average of all hillslope vectors was taken for each convergent and divergent sample in each study area to provide a mean hillslope vector. The average length of the ten hillslopes was used to determine the average hillslope length.

Data files were set up using the Splus language (Statistical Sciences 1993) for graphic creation. The E horizon probability data was used to filter the E horizon depth so that E horizon depths are only displayed for those hillslope distance vectors where the probability of an E horizon is greater than 0.5. Approximate colors were used that closely match the true soil colors as determined from the sample data. The graphical axes (x = hillslope distance, y= hillslope height) were established to fit all of the convergent or divergent hillslope model displays to highlight relative differences. The layer depths have been multiplied by a factor of ten to enhance cross-section display, and the ordinate extended to negative hillslope heights to fit the entire soil layer extents onto the display.

3. Results and Discussion

Figure 2 displays the convergent hillslope models and Figure 3 the divergent hillslope models. The relative differences in basic hillslope structure for each study area are readily apparent from the graphics. The soil layers may now be used to interpret and compare how these landscapes function with respect to material and energy pathways and flows.

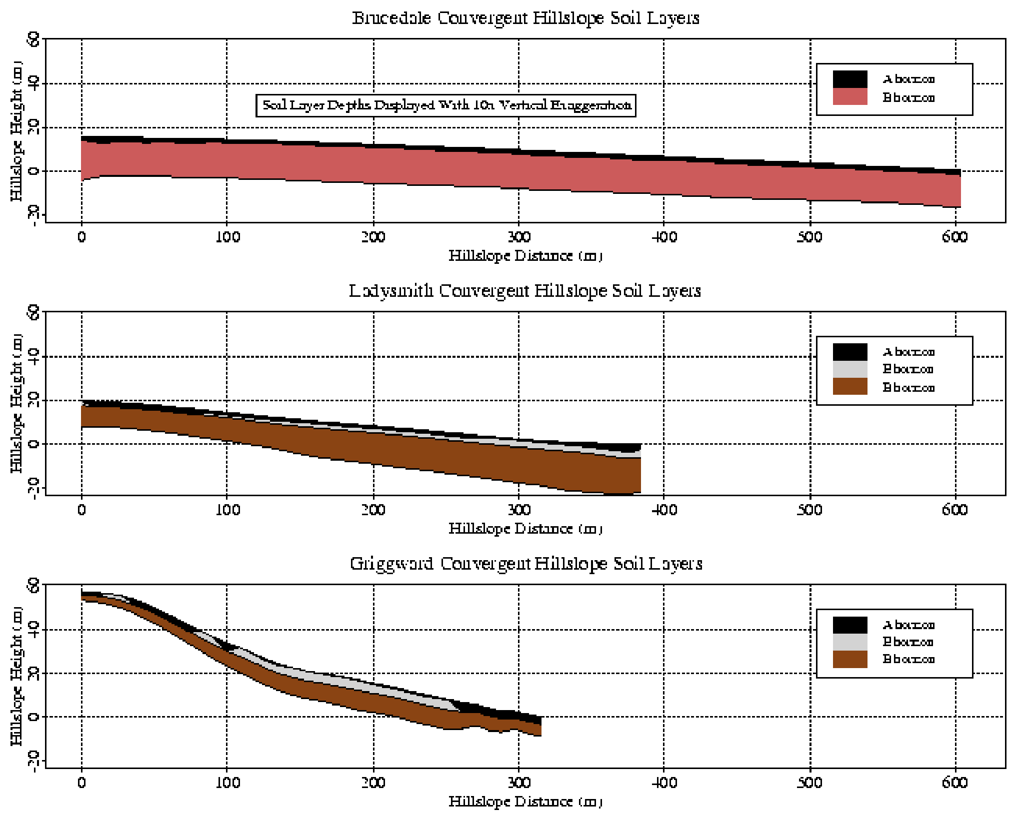


Figure 2. Study Area Mean Convergent Hillslope Models

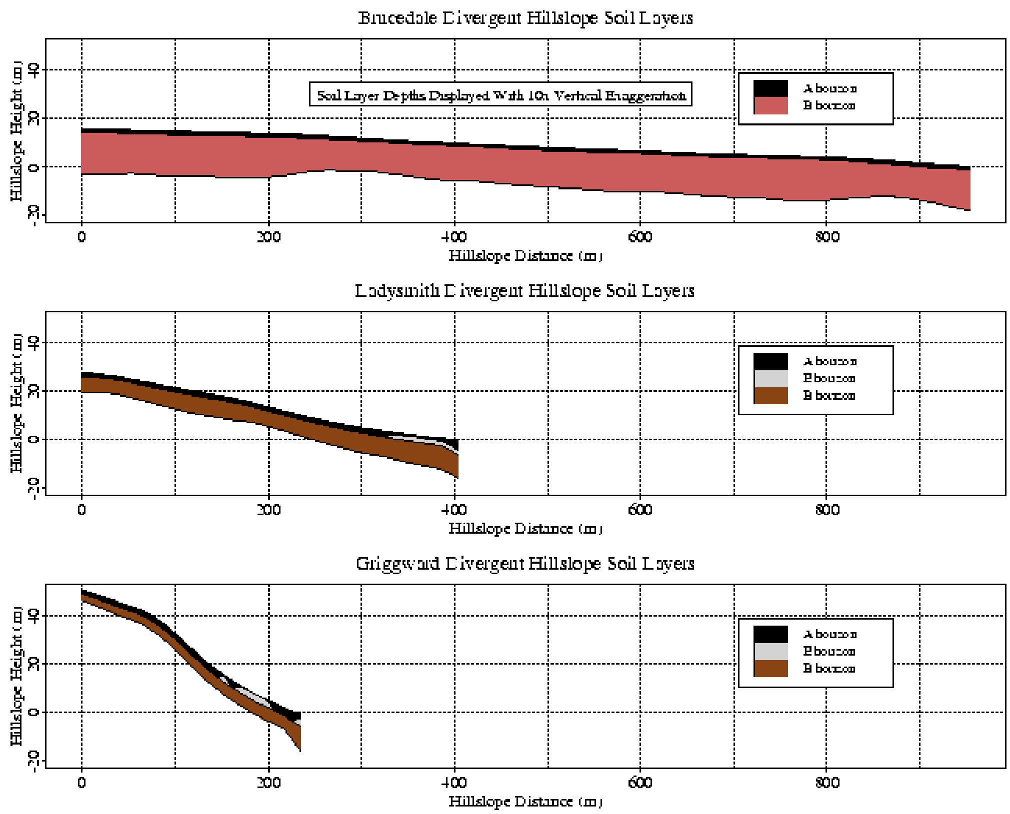


Figure 3. Study Area Mean Divergent Hillslope Models

3.1 Brucedale Hillslope Models

The A horizon depths are slightly greater and the solum depth less variable on the Brucedale mean convergent hillslope, but the overall difference between convergent and divergent hillslopes is small. The consistent nature of the A horizon depth may be influenced by the widespread cereal cropping landuse in this study area where cultivation maintains a plow layer A horizon (Ap). Soil texture does not strongly contrast between the A horizon (clay loam) and B horizon (light clay). Field samples also indicated that the B horizons are well structured. These factors coupled with the dryer and warmer climate and gently inclining topography indicate that very little surface or subsurface lateral flow occurs in these landscapes. This suggests that the principal material and energy flow pathway in this landscape is in-situ vertical infiltration of water where it is either used by plants or lost to deep percolation.

3.2 Ladysmith and Griggward Hillslope Models

The A horizon depths are deeper and more variable in the Ordovician metasediment landscapes. The A horizon depths for Griggward are deeper than Ladysmith and occupy a larger proportion of the solum. The soil texture contrasts more markedly in the solum derived from the Ordovician metasediments parent materials compared to the Brucedale setting. The A horizons are typically sandy loams and the B horizons, light clay to clay textures. This texture contrast provides greater impedance to vertical conductance of water from the A to B horizons. This factor coupled with the steeper slopes and relatively cooler and moister conditions compared with Brucedale favors development of distinct E horizons where lateral throughflow of water is an important process.

As expected, the E horizons occur much higher on the convergent hillslopes for both study areas with the Ladysmith landscapes showing a more systematic E horizon development and increase in depth transcending the hillslope. Both convergent and divergent hillslopes for Griggward show a more probabilistically variable occurrence of E horizons. This suggests that the moister, cooler and higher energy Griggward landscapes exhibit a more dynamic mixture of processes including overland flow, material movement by surface erosion and subsurface flow.

Lack of E horizons just below the hillslope summit on the Griggward convergent hillslope may indicate an area of active colluvial processes not stable enough for E horizon development to occur. E horizon absence at the hillslope base may indicate an active area of alluvial processes as the hillslope connects to the broader, higher energy watershed system. The area in between where deep E horizons occur may be a region of seasonal waterlogging.

3.3 Regional Interpretations

Over the climatic gradient in the direction of increasing rainfall and cooler temperatures, A horizon depths increase and solum depths decrease. This is echoed by an increase in the A horizon mean total carbon percentages (Brucedale: 1.3, Ladysmith: 2.27; Griggward: 2.58) indicating increased soil biological activity. Hence, a more favorable annual water balance, less prevalent dry and hot organic matter oxidizing conditions and less frequent soil cultivation likely contribute to this. The Brucedale soils exhibit slightly higher cation exchange capacity's and higher pH's than the Ladysmith and Griggward soils. This suggests conditions more favorable to leaching and solute transport in the Ordovician metasediment study areas reflected by the E horizon presence. The descreased solum depths on the Ordovician metasediments suggest that erosional processes are more common as a result of soil layer, geomorphometric and climatic conditions. These indicate that lateral surface and subsurface flow pathways are much more prevalent in the Ordovician metasediment landscapes.

4. Conclusions

This paper presents a summary of our progress in soil-landscape modelling research and briefly demonstrates new work that integrates developing methods. The intention is to demonstrate quantitative methods that place soil-landscape modelling on a scientific foundation for continual evaluation and improvement. In many hydrological and environmental models, the soil component is incorporated in a very simplistic way, if at all. The methods discussed here aim to provide continuous spatial predictions with known levels of confidence based on the sample evidence collected. The integration of models to develop hillslope soil layer visualizations is simple in concept, yet provides an explicit and quantitative base from which to build better models and understanding of landscape processes and dynamics. The models behind the visualizations may also be used to rapidly generate automated maps for landscape management.

4.1 Critical Appraisal of Approach

Several general areas are in need of further conceptual development and empirical testing. At the broad regional scale, we have developed and characterized three study areas using prior information about geology and climate. More research is required to determine how to best define physiographic domains and place them in a broader spatial context to extrapolate developed models. At the more local catchment and hillslope scales, we have used a simple environmental gradient, based on the compound topographic index terrain attribute, to allocate and spread samples in both environmental attribute and geographic space. More research is required to define basic concepts to guide this process. These are difficult challenges for general purpose natural resource inventory intended to build various models of different attributes from a single sampling. Iterative and nested approaches for sampling and data analysis may be more appropriate, but are not always logistically feasible.

We have maintained information that could be used to account for and model error propagation through the modelling process, but we have not attempted to explicitly incorporate it. It is likely that this should be planned from the beginning. Tools to do this need to be developed.

5. Acknowledgements

The authors thank Linda Ashton, Deborah O'Connell, Jeff Wood and Dermot McKane for assistance. This study was funded in part by Grant No. NRMS-M218 from the Murray-Darling Basin Commission.

6. References

Akaike, H. (1974) A new look at statistical model identification. *IEEE Transactions on Automatic Control*, AU-19:716-722.

Breiman, L., Friedman, J.H., Olshen, R. and Stone, C.J. (1984) *Classification and regression trees*. (Belmont, California:

Wadsworth International Group).

Buol, S.W., Hole, F.D. and McCracken, R.J. (1989) *Soil genesis and classification*. Third Edition. (Ames: Iowa State Univ. Press).

Carson, M.A. and Kirkby, J.J. (1972) *Hillslope form and processes*. (Cambridge: Cambridge University Press).

Clark, L.A. and Pregibon, D. (1992) Tree based models. In *Statistical Models in S*. (Pacific Grove, California: Wadsworth & Brooks), pp. 377-419.

Forman, R.T.T., and Godron, M. (1986) *Landscape ecology*. New York: Wiley.

Geeves, G.W., Cresswell, H.P., Murphy, B.W., Gessler, P.E., Chartres, C.J., Little, I.P. and Bowman, G.M. (1995) The physical, chemical and morphological properties of soils in the wheat-belt of southern N.S.W. and northern Victoria. *CSIRO Divisional Report X*. (CSIRO Div. of Soils: Canberra), pp. 178.

Gerrard, A.J. (1981) *Soils and landforms: an integration of geomorphology and pedology*. (London: Allen & Unwin).

Gessler, P.E., Moore, I.D., McKenzie, N.J. and Ryan, P.J. (1995) Soil-landscape modelling and spatial prediction of soil attributes. *Int. J. Geographical Information Systems*, Vol. 9, 4:421-432.

Hastie, T. and R. Tibshirani. 1990. *Generalized additive models*. (Chapman & Hall: New York).

McCullagh, P. and J.A. Nelder. 1989. *Generalized linear models*. (Second Edition. Chapman and Hall: London).

McKenzie, N.J. and Austin, M.P. (1993) A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. *Geoderma*, 57:329-355.

McMahon, J.P., Hutchinson, M.F., Nix, H.A. and Ord, K.D. (1995) ANUCLIM: Users guide. Canberra: Centre for Resource & Environmental Studies, Australian National University.

McSweeney, K.M., Gessler, P.E., Slater, B., Hammer, R.D., Bell, J. and Peterson, G.W. (1994) Towards a new framework for modelling the soil-landscape continuum. In *Factors of Soil Formation: A Fiftieth Anniversary Retrospective*, Special Publication 33 (Madison, WI: Soil Science Society of America), 127-145.

Moore, I.D., Gessler, P.E., Nielsen, G.A. and Petersen, G.A. (1993) Soil attribute prediction using terrain analysis. *Soil Science Society America Journal*, 57:443-452.

Turner, M.G. and Gardner, R.H. (1991) Quantitative methods in landscape ecology: an introduction. In *Quantitative Methods in Landscape Ecology*. Turner and Gardner (eds.) Ecological Studies 82. (New York: Springer Verlag) pp. 3-14.

Walker, P.H. and Butler, B.E. (1983) Fluvial processes. In *Soils: an Australian viewpoint*. Division of Soils, CSIRO (Melbourne: CSIRO/Academic Press) pp.83-90.

Paul Gessler and Neil McKenzie
CSIRO Division of Soils
GPO Box 639
Canberra, ACT 2601
Australia
Telephone: +61 6 246 5942
Fax: +61 6 246 5965

Michael Hutchinson
Centre for Resource and Environmental Studies
Institute of Advanced Studies
Australian National University
Canberra, ACT 0200
Australia
Fax: +61 6 249 0757

Bheshem Ramlal and Kate Beard

An Alternate Paradigm for Representing Soils Data and Data Quality Information

---

Abstract

The use of the traditional map unit based model for representing soils data and its accompanying data quality information suffers many limitations. The primary assumption that the soil landscape consists of discret e, nearly homogeneous units, is not generally valid. This model limits the effective storage, management and use of soils data and data quality information in the GIS environment.

This paper proposes the use of an alternate model that views the soil landscape as being predominantly continuous in nature and occasionally interrupted by abrupt changes. This model of reality is called the mixed variation model. To demonstrate the advantages gained from this paradigm shift, a conceptual model for a measur ement based soil information system that uses this assumption is presented. In this model point observations on soil properties, ground penetrating radar data, and change delineation are stored along with associated quality descriptors. This paper describ es the model, the comparative analysis carried out to evaluate it and presents a discussion of the results. Special data and data processing requirements of this model are described and the benefits of using this model are discussed.

Introducti on

> *No amount of data storage and retrieval technology can compensate for an inappropriate conceptual model of soil variation.* (Burrough 1993, pg. 19)

There are primarily two paradigms that have been use d to conceptualize geographic space. The first represents space as objects with well defined boundaries and attributes. The second represents geographic reality as a continuous field of data (Burrough and Frank 1995, Couclelis 1992). Although certain type s of data in the geographic space may be effectively represented with the use of the object view of reality, it is not the most appropriate model for other types of natural phenomenon such as soil landscape properties. Although the inadequacy of using thi s model for soil information has been conceded for some time (Hole and Campbell 1985), its use persists. This may be attributed to the fact that this model has been used for more than a century of soil mapping (Soil Survey Staff 1993), and up until the la st decade, the technology did not exist to allow the use of any other paradigm for modeling this landscape (Ernstrom and Lytle 1993).

The continuous field model of reality is appropriate for use with data that vary continuously over space. It may be used to represent the soil landscape. Although it provides a better representation than the object model of the continuous nature of most soil properties, it does not accommodate sharp changes that may occur. There is therefore a need for an alternativ e model that manages both types of spatial variation.

If we assume that the object model and the continuous field model are the two extremes of a continuum, then a model that is capable of handling both types of data may lie between these two mod els. We call this the mixed variation model. These three models are shown in figure 1.

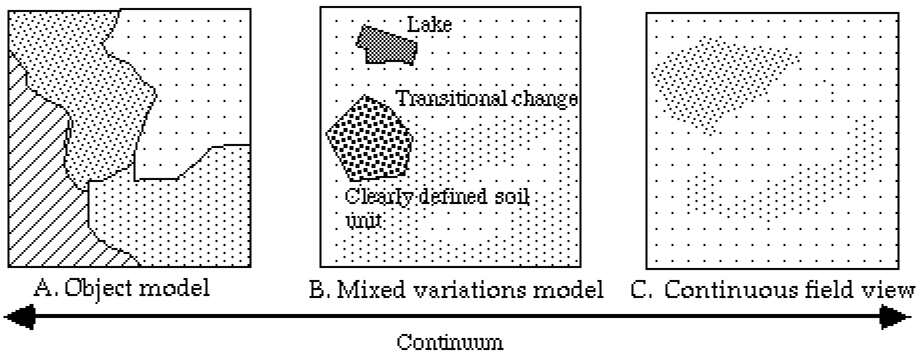


Figure 1. Models of Reality

Diverse models of reality force views of the landscape through different filters. To represent these particular views on maps o r in the GIS environment, sampling schemes are designed so that the data needed to recreate these representations may be acquired. Many assumptions are made in the design of these sampling procedures. Once data have been acquired for a particular view, it is difficult, if not impossible to use these data to create a different view of reality. If an inappropriate model of reality is used, we are not only left with an inadequate representation but with data that may not be very useful for anything else. It may be noted from figure 2. below, that choosing the wrong model limits our ability to move to another. For example, it is not possible to use data from an object model representation to create a mixed variations representation. However, the reverse is po ssible. It is also noted that changing from one model to the next is sometimes unnecessary or even meaningless. Problems that arise in using an inappropriate model are the possible limits placed on our ability to model error and assess the quality of data . These problems are discussed below.

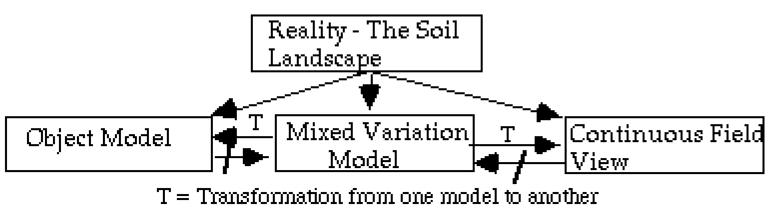


Figure 2. Choosing the appropriate model of reality

This paper will highlight the major limitations of using the object model for representing the soil landscape and for providing data quality information for this representation . An alternative model that views the landscape as comprised of mixed variations is forwarded as an alternative to address these limitations. The advantages of using this alternate paradigm are presented. Special data and data processing requirements of t his model are described. Finally, the benefits gained are discussed. It should be noted that the continuous field model is not discussed further although it has been used for representing individual soil properties (see for example Yost et al. 1982 a,b, a nd McBratney et al. 1992). We argue that it has limitations because of its inability to accommodate sharp changes in soil properties.

Before discussing the limitations of the object model for representing the soil landscape, a brief description o f the soil landscape is given. This provides an indication of the complex variation that must be captured and represented in the spatial information system. Soils consist of numerous chemical, physical and biological properties that vary in space and time . These variations in properties, while not totally independent of each other, are neither uniform nor abrupt for the most part; instead, changes are normally transitional, with some exceptions of abrupt changes. Variations are apparent at diverse scales. Patterns are discernible at a continuum of scales from as little as a few meters to many kilometers (Trangmar et al. 1985). Additionally, variations are not just two-dimensional; soil properties may significantly change with depth and time.

Limi tations of the Object Model

The object model breaks up the complex nature of the soil landscape into mapping units that are viewed as being internally homogeneous and demarcated by sharp boundaries (Hole and Campbell 1985). Each of these units is assigned a soil class and a representative profile. A soil class is established by designating the ranges of values for five to seven soil properties that the soil must satisfy. A representative profile for a soil class shows the structure of the profile and ranges of soil property values that are expected to be found for a given soil class (Soil Survey Staff 1993, 1994).

The above method of viewing the soil landscape significantly influences the strategies used to map it. In practice, data are ac quired in a two stage process. The first stage acquires data that are used to generate representative profiles and soil classes. The second stage involves the delineation of soil mapping units on aerial photographs and the assignment of a class to each of these units (Valentine 1986).

The resulting soil survey provides a soil map with very little point specific data. Additionally, these data undergo significant processing and generalization. As a consequence, the object model limits the data that are finally stored in the database. This mapping process leads to a loss of data that cannot be recovered(Zinck 1993), including original observations taken in the field and the means to model errors. This makes it increasingly difficult to provide measu res that are appropriate in the determination of the quality of these data (Burrough 1993).

A major problem caused by this model is the separation of error into positional and attribute accuracy. Splitting these components for independent determi nation is not very meaningful since these are highly correlated (Chrisman 1989, Veregin 1989). Additional problems in dealing with other quality components arise. For example, information on the completeness or lineage for individual soil units in the dat aset may be difficult to obtain. Information on the effects of processing steps are also not available. In a GIS environment where data may be easily shared and combined, an absence of data quality information may have serious consequences. Reliability in formation is at best only available at the class level and information on the reliability of individual properties is unavailable. Mapping units may in fact mask the spatial variation in certain individual soil properties.

To highlight the major limitations of the object model, we use an object model based schema forwarded by Fernandez et al. (1993 a,b) and modify it to include data quality information. The inclusion of this information is based on the assumption that purity measures, as defined by Marsman and de Gruijter (1986), are available. These measures give accuracy values for the soil class, the accuracy of the individual properties used in the classification and the average accuracy of these properties. Purity measures are obtained by co mparing the soil type on the map with what is found in the field (Marsman and de Gruijter 1986). The modified schema is shown in figure 3.

Limitations are identified by analyzing how the object model provides data and data quality information for a single soil property for the soil landscape. What is significant is the capability to provide a usable quality measure. We assumed that this database was implemented and populated with data and data quality information. We imagined a scenario where a u ser is interested in obtaining a thematic surface map and a reliability map of the soil texture for the A- horizon for an area of interest. The way this database provides this information is described below.

The soil database is searched for the a ssigned soil type for each soil mapping unit. Attached to each of these soil types is a representative profile. Each profile consists of layers or horizons. Data about the physical characteristics are stored for each layer. From these layer data, a range of values for soil texture is extracted for the A horizon and assigned to each soil mapping unit. An average value of the range is then calculated and reassigned to each unit. The purity measure which applies to soil classes is retrieved from the data qua lity information and assigned to each mapping unit. The information returned will be in the form of two thematic maps. The first one shows average texture values for soil mapping units. The second map shows purity values for these texture values.

Both the soil information and the data quality information from this system are limited. The surface map contains homogeneous units of soil texture values. These values were not actually measured within the soil unit, but were extrapolated from represent ative profiles of different soil types and assigned to these units. While this may sometimes give a good estimate, there is no way of knowing whether these values will be found within these mapping units. An additional problem is that these values are ave rages obtained from ranges that were estimated in the field for a few profiles. The reliability map is limited in that purity estimates are done for soil types. As a consequence, these estimates do not reflect the reliability for individual soil units or properties but representative units. Additionally, these purity estimates were taken for the ranges of values not for single values. This makes it difficult to derive useful quality measures.

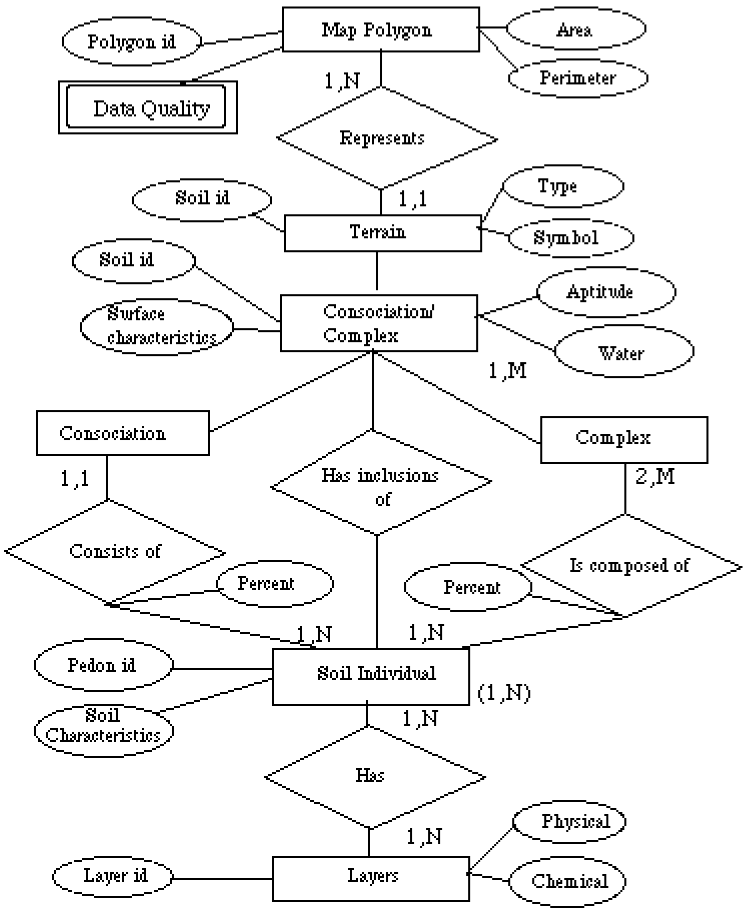


Figure 3. Conceptual Schema for an Object Model Based Soil Information System (after Fernandez et al. 1993 a,b).Modified to include data quality.

The Mixed Variation Approach

Given the shortcomings of using the object model for representing the soil landscape and for modeling errors, we propose an alternative system that preserves data in a form that allows the recreation of a closer representation of the land scape and allows the assessment of the uncertainty of this representation. The overall differences in the way this system are expected to handle soil data are compared in figure 4.

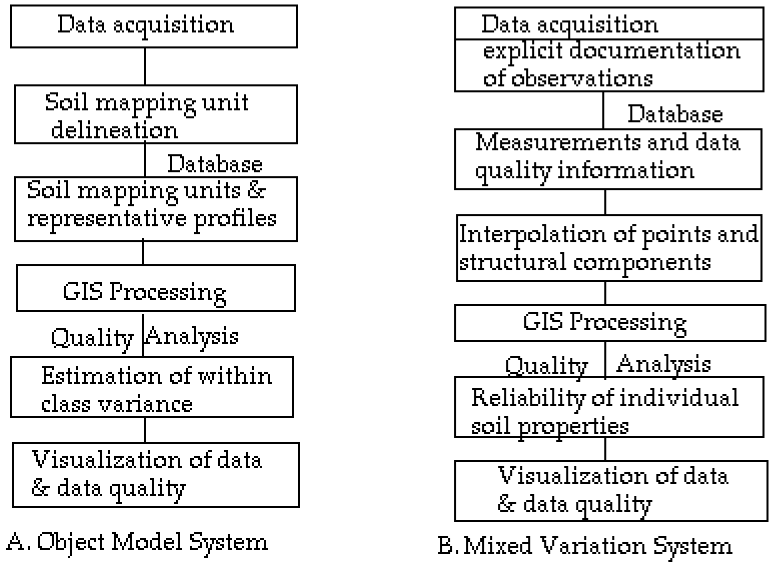


Figure 4. A comparison of object model and mixed variation based systems

The fundamental difference betw een the object and the mixed variation models lies in the data stored and the strategies used in the acquisition and processing of these data. This system will store at one level, two data types: (1) raw observations taken mostly at points and (2) abrupt changes that occur along lines (See figure 5). The data are expected to come from diverse sources: field surveys, aerial photography, global positioning systems (GPS), and ground penetrating radar (GPR). Unlike the traditional technique of recording only a few profiles subjectively chosen, this model requires systematic sampling. Instead of using a few properties to classify a soil in the field, this method will measure, observe and record specific soil properties at all points visited where possible. Sha rp changes in soil properties observed in the field and on aerial photographs are also recorded. The delineation of sharp changes may require expert knowledge. As a consequence, experience is not wasted in the new system.

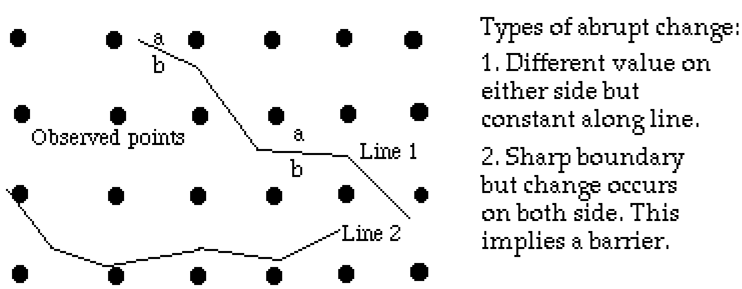


Figure 5. Two data types of the mixed variation model: points and lines of abrupt change.

A significant departure from the traditional technique is the storage of data quality information with individual measurements in the database. Information on the data source, method and date of measurement, calibration data, accuracy, resolutio n, completeness, and consistency of all measurements are to be included in the database (see figure 6). This information is useful for determining the fitness for use of different measurements stored in the database.

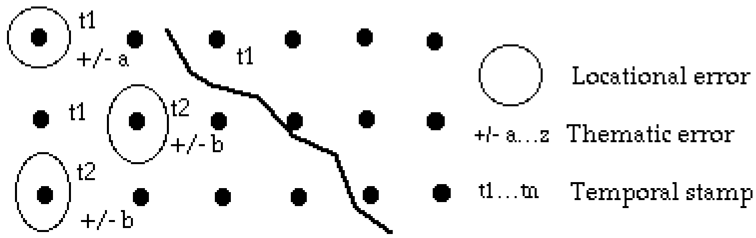


Figure 6. Mixed variation data that vary in locational, thematic and temporal accuracies

We designed a conceptual schema for a mixed variation model database that stores both soils data and data quality information (see figure 7). This design incorporates data from different sources that vary in quality. Note that data quality information is stored at different levels of detail. However, the same information is not stored twice. The lower level entities inherit quality information from higher level entities.

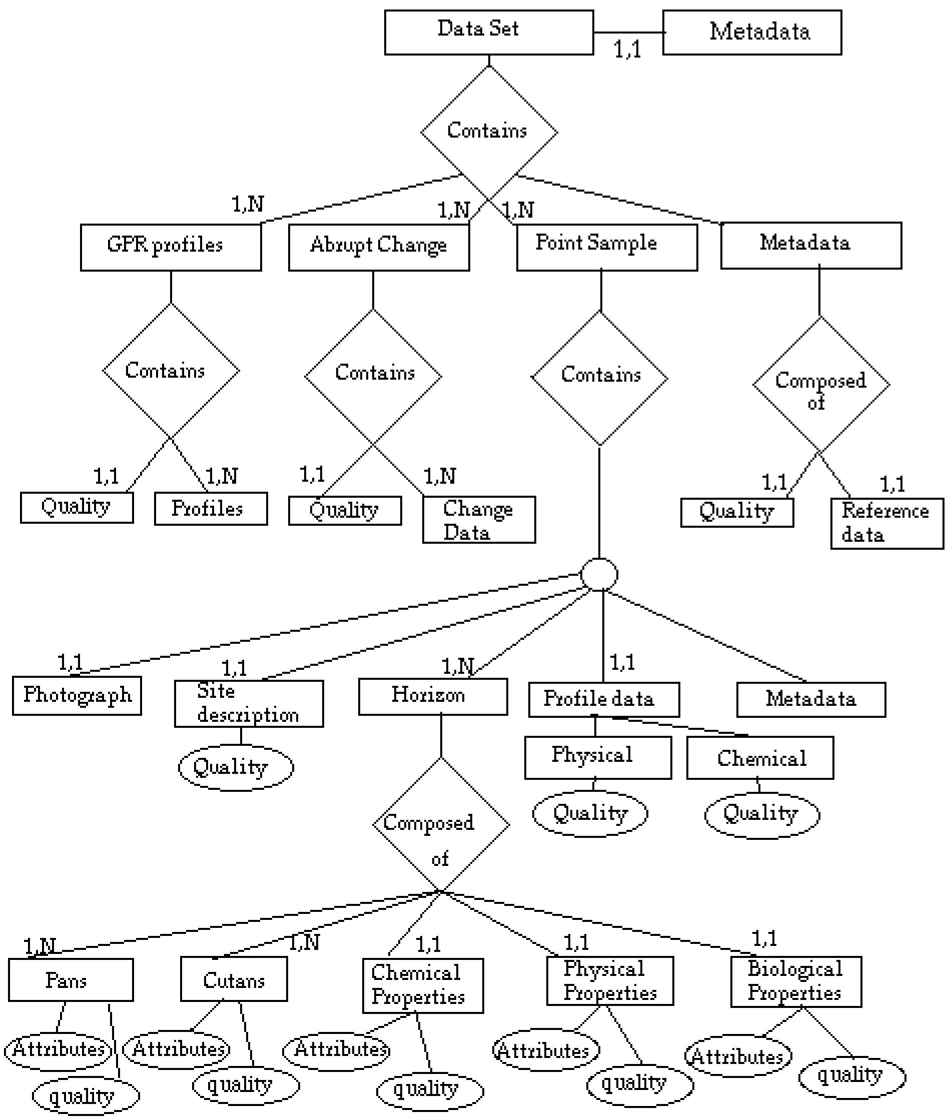


Figure 7. A conceptual model for a mixed variation soil information system. This model includes data quality information

The same scenario used for highlighting the limitations of the object model is used here to describe how the mixed variation model will respond to a query for information. The following procedure is expected to be used to generate the required information. The user specifies a resolution for t he information requested. The system may have the intelligence to give a resolution range that is appropriate. Once the resolution is chosen, the system will then search the database to identify the data available for satisfying this request. Point sample s are examined first. This involves searching down the hierarchy to the physical properties for horizons and extracting a texture value for each sample point. Then data from the dataset of abrupt change delineation are extracted based on the relevant reso lution and soil property. The attached data quality information must be used to identify these data. Next, the GPR profiles are checked for data that satisfy the resolution and soil property requirements. Resampling of GPR profiles may be necessary to avo id skewing the results in one direction. Once the data are obtained, it is necessary to assign weights to them. Weighting values may be determined using the accuracy of the soil texture measurements. Interpolation is then used to generate the required sur face map.

The mixed variation model based system provides data that are reflective of the soil landscape. The surface is generated from actual measurements. The system takes into consideration that these data may have come from different sources w ith different accuracy levels and weighs them accordingly. Also significant are the data about sharp changes in soil properties. This combination accounts for both transitional and non-transitional changes that may occur. We note the importance of data qu ality. This information is not only important to filter out the unwanted data but it also provides the information for assigning weights. Additionally, this information is used to generate the reliability map. This map reflects the reliability of data der ived from measurements. Its generation is easier to achieve than in the previous system. Another benefit of storing data quality information is its use to generate quality measures prior to data processing. This saves time and money (Burrough 1992).

Special Requirements of the Mixed Variation Based Soil Information System

The shift in paradigm from a object model to a mixed variation based model changes the requirements for data acquisition, data storage, data processing, and error mode ling.

In the mixed variation model rather than ascertaining that a soil belongs to a set of classes, soil properties are measured or observed, and recorded. The resulting data consist of qualitative and quantitative measures of soil properties. GP S surveys, satellite imagery and aerial photography are sources for abrupt change information. Experienced soil surveyors may delineate these changes on rectified images or using GPS receivers in the field. Because changes may be considered sharp at diffe rent resolutions for different soil properties, the storage of data quality information for these data are especially important. The soil property for which the change applies, the relevant resolution, the accuracy of delineation, and source of the data a re examples of the quality information stored with abrupt change objects. GPR data consist of cross-sectional profiles taken at intervals on the landscape. Along the transects of these profiles, the coverage is continuous. However, these transects may be spaced from a few meters to hundreds of meters apart. This results in a skewed coverage of the soil landscape. The depth resolution of the GPR profile may also vary depending on the material that reflects the signal back to the sensor (Doolittle 1987). Th ese factors must be taken into consideration when these data are used.

The strategies used for sampling data need to be revisited. Instead of a subjective model of sampling that uses mostly expert knowledge to delineate soil boundaries and estimat e soil property values, a systematic approach that acquires quantitative data, is required. Sampling strategies that will provide optimal results using interpolation techniques are needed. Some soil scientists have suggested that regular grid surveys are the most appropriate (Burgess and Webster 1980a). Others have argued that the use of equilateral grids with some sampling at smaller intervals within the grid is most appropriate (Trangmar et al. 1985). The use of random sampling has also been proposed (V an Kuilenberg et al. 1982). A solution has been forwarded by McBratney and Webster (1983a) for choosing an appropriate method of sampling. They suggest the use of preliminary sampling to obtain a working semi-variogram from which a method for optimal samp ling may be designed for best interpolation results. See Trangmar et al. (1985) and Cressie (1991) for a discussion on semi-variograms and their use in geostatistics.

The mixed variation model is based on the storage of measurements. The success o f the proposed system is centered on the use of interpolation techniques for processing these data. Comparative studies have shown that kriging is the optimal method for interpolating point data (Van Kuilenberg et al. 1982, McBratney and Webster 1983b, Bu rgess et al. 1981, Webster and Oliver 1989, Webster and Burgess 1980, Burgess and Webster 1980 a,b, Goovaerts 1992, Voltz and Webster 1990). An important consideration for the proposed system is the processing of both point data and lines delineating disc ontinuities in the landscape. The use of kriging with sharp delineations has been employed by a few soil scientists (Stein 1994, and Heuvelink and Bierkens 1992). However, only enclosed mapping units were considered. A method to accommodate linear discont inuities is required. An additional requirement that must be satisfied is the handling of measurements that vary in accuracy. Rather than exact interpolation to sample measurements, techniques are required that will change these values based on confidence limits. This is the subject of a future paper.

Benefits of a Mixed Variation Based Soil Information System

The mixed variation approach offers a number of benefits. Soil data need not be left out because of processing, classification or an inability to handle large amounts of data (Ernstrom and Lytle 1993). The soil landscape are better represented with this approach since both the continuous nature as well as abrupt changes that may be present, are integrated into the database for stora ge, retrieval and manipulation. Very specific information may be generated using this system. It is also possible to use the database to generate traditional mapping units in the form of representation of individual soil properties, soil classes and repre sentative mapping units if desired.

This approach provides information that has so far been omitted or cannot be obtained using the object model: data quality information. What is different as well is that this information is obtainable for indivi dual measurements, sets of measurements, or for the entire data set. Additionally, data quality may be generated for single soil properties and interpolation results of individual properties rather than soil classes.

Conclusions

The propo sed system offers many advantages over the traditional object model based system for representing the soil landscape and for providing data that allows the assessment of the uncertainty of this representation. Original, unprocessed data are available to u sers to be processed into whatever form they may choose. While there are many advantages to this system, these are only gained at some cost. It is expected that costs of data acquisition will rise significantly. However, further development in technology will bring down the cost of field sampling and laboratory analysis.

References

Bierkens, M.F.P. and Burrough, P.A. (1993a). The Indicator approach to categorical soil data I. Theory. *Journal of Soil Science* 44: 361-368.

Bier kens, M.F.P. and Burrough, P.A. (1993b). The indicator approach to categorical soil data II. Application to mapping and land use suitability analysis. *Journal of Soil Science* 44: 369-381.

Burgess, T.M. Webster, R. and McBratney, A.B. (1981) Optimal and isarithmic mapping of soil properties: IV. Sampling Strategy. *Journal of Soil Science* 32:643-659.

Burgess, T.M. and Webster, R. (1980a). Optimal interpolation and isarithmic mapping of soil properties I. The semi-variogram and punc tual Kriging. *Journal of Soil Science* 31: 315-331.

Burgess, T.M. and Webster, R. (1980b). Optimal interpolation and isarithmic mapping of soil properties II. Block Kriging. *Journal of Soil Science* 31: 333-341.

Burrough, P.A. (1 992). Development of intelligent geographic information systems. *IJGIS* 6.1: 1-11. ITC Journal 1993-1: 15-22.

Burrough, P.A. and Frank, A.U. (1995). Concepts and paradigms in spatial information: are current geographical information systems truly generic? *IJGIS* 9.2: 101-116.

Cressie, N. (1991) *Statistics for spatial data*, Wiley, New York ,

Doolittle, J.A. (1987). *Using Ground-penetrating Radar to Increase the Quality and Efficiency of Soil Surveys. Soil Survey Techniques.* SSSA Inc. 98.

Ernstrom, D.J. and Lytle, D. (1993). Enhanced soil information systems from advanc es in computer technology. *Geoderma* 60: 327-341.

Fernandez, R. N. and Rusinkiewicz, M. (1993a). A conceptual design of a soil database for a geographical information system. *IJGIS* 7.6: 525-539.

Fernandez, R.N. Rusinkiewicz, M. da Silva, L. M. Johannsen, C.J. (1993b). Design and implementation of a soil geographic database for rural planning management. *Journal of Soil and Water Conservation* March-April 1993: 140-146.

Goovaerts, P. (1992). Factorial Kriging analysis : a useful tool for exploring the structure of multivariate spatial soil information. *Journal of Soil Science* 43: 597-619.

Heuvelink, G.B.M. and Bierkens, M.F.P. (1992). Combining soil maps with interpolations from point observations to pred ict quantitative soil properties. *Geoderma* 55: 1-15.

Hole, F.D. and Campbell, J.B. (1985). *Soil Landscape Analysis.* New Jersey: Rowman and Allanheld.

Marsman, B.A. and de Gruijter, J.J. (1986). *Quality of soil maps*. Soil Survey Institute, Wageningen, The Netherlands.

McBratney, A.B. De Gruijter, J.J. Brus, D.J. (1992). Spacial prediction and mapping of continuous soil classes. *Geoderma* 54: 39-64.

McBratney, A.B. and Webster, R. (1983a). Optimal i nterpolation and isarithmic mapping of soil properties V. Co-regionalization and multiple sampling strategy. *Journal of Soil Science* 34: 137-162.

McBratney, A.B. and Webster, R. (1983b). How many observations are needed for regional estima tion of soil properties. *Journal of Soil Science* 38.3: 177-183.

Soil Survey Staff (1993). *Soil Survey Manual. USDA Handbook No.18.* US Department of Agriculture.

Soil Survey Staff (1994). *National Soil Survey Handbook.* USDA, Soil Conservation Service.

Stein, A. (1994). The use of prior information in spatial statistics. *Geoderma* 62: 199-216.

Stein, A. Hoogerwerf, M. Bouma, J. (1988). Use of soil-map delineations to improve (co-)kriging of point data on moisture deficits. *Geoderma* 43: 163-177.

Trangmar, B.B. Yost, R.S. Uehara, G. (1985). Application of geostatistics to spatial studies of soil properties. *Advances in Agronomy*. Ed. N.C. Brady. Orlando: Academic Press Inc. 38: 45-94.

Valentine, K.W.G. (1986) *Soil Resource Surveys for Forestry : Soil, terrain, and site mapping in boreal and temperate forests.* Oxford Science Publications, New York.

Van Kuilenburg, J. De Gruijter, J.J. Marsman, B.A. Bouma, J. (1982). Accuracy of spatial interpolation between point data on soil moisture supply capacity, compared with estimates from mapping units. *Geoderma* 27: 311-325.

Veregin, H. (1989). *A Taxonomy of Error in Spatial Databases.* NCGIA. 8 9-12.

Voltz, M. and Webster, R. (1990). A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. *Journal of Soil Science* 41: 473-490.

Webster, R. and Burgess, T.M. (1980). Optimal interpolation and isarithmic mapping of soil properties I. Changing drift and universal Kriging. *Journal of Soil Science* 31: 505-524.

Webster, R. and Oliver, M.A. (1989). Optimal interpolation and isarithmic mapping of soil properties VI. Disjunctive Kriging and mapping the conditional probability. *Journal of Soil Science* 40: 497-512.

Yost, R.S. Uehara, G. Fox, R.L. (1982). Geostatistical analysis of soil chemical properties of large land area. I. Semi-variograms. *Soil Sc i. Soc. Am. J.* 46: 1028-1032.

Yost, R.S. Uehara, G. Fox, R.L. (1982). Geostatistical analysis of soil chemical properties of large land area. II. Kriging. *Soil Sci. Soc. Am. J.* 46: 1033-1037.

Zinck, A.J. (1993). Introduction. *ITC Journal 1993-1. Special Issue on Soil Survey Workshop*: 2-7.

Bheshem Ramlal, Graduate Student, Department of Spatial Information Science and Engineering, University of Maine, 346 Boardman Hall, Orono, ME 04469-5711, bramlal@spatial.maine.edu

Kate Beard, Associate Professor, NCGIA and Department of Spatial Information Science and Engineering, University of Maine, 346 Boardman Hall, Orono, ME 04469-5711, beard@spatial.maine.edu

Andrew S. Rogowski, and Edwin T. Engman

USING A SAR IMAGE AND A DECISION SUPPORT SYSTEM TO MODEL SPATIAL DISTRIBUTION OF SOIL WATER IN A GIS FRAMEWORK

ABSTRACT

Remotely sensed images of soil moisture obtained with a synthetic aperture radar (SAR) contrast sharply with point measurements taken on the ground. Point measurements of soil water represent a single realization of soil water distribution at a given time, while a remotely sensed image is a digital equivalent of a spatially continuous scene. The study area is generally characterized by spatially varying soil properties, multiple landforms, and different vegetation sequences. Local values are not independent, but rely on the distribution of incoming precipitation, antecedent soil water conditions, history of moisture extraction by roots, and relative position within a landscape. This type of information is indirectly imbedded in a SAR image which represents a continuous soil water distribution, averaged over a pixel area at a given scale of resolution. We make use of a *geostatistical simulated* annealing approach to construct synthetic GIS overlays of surface soil water based on measured values. Soil water contents measured at a point in the top 5 cm are utilized to estimate changes in soil water levels over larger blocks at a 3×3 SAR pixel scale of resolution. The synthetic SAR images are adjusted for the effects of soil type, slope, aspect and land use. To make adjustments, we apply a multiple-criteria decision-making procedure (MCDM) known as *compromise programming* within a GIS framework. Changes in synthetic images are compared to corresponding changes in SAR scenes. Results illustrate how to generate statistically equivalent images of remotely sensed data and how to adjust them for random effects of terrain attributes. The suggested approach quantifies the correspondence between soil water changes in the field and perceived changes in backscatter intensity for remotely sensed SAR scenes. One weakness is that in this study the approach relies on expert knowledge to choose approriate weights. However, since the sensitivity of radar backscatter to the above terrain attributes is well documented in literature, the rather arbitrary choice of weights will eventually be replaced by more objective deterministic, or stochastic algorithms now under development.

INTRODUCTION

The hydrologic cycle's interaction with the Earth's land surface occurs within a thin reservoir that stores and distributes (spatially and temporally) water that falls on the surface in the form of rain or melting snow. This reservoir is commonly referred to as *soil water*. Soil water is an environmental descriptor that integrates much of the land surface hydrology, acts as the interface between the land surface and the atmosphere, and plays a pivotal role in life's biochemical and geochemical cycles. In spite of its obvious importance, soil water has not had wide spread use as a variable in land process models. There are two primary reasons for this. First of all, soil water means different things to different people and there is no consensus on the observational requirements (i.e., depth, accuracy, frequency and spatial representation). For example, a mesoscale modeler might want two layers of soil water measurements (0-5 cm and 5-100 cm) on a grid scale of 30 km whereas a hydrologic modeler may be content with a single 10 cm layer of soil water but at a spatial scale of 30 m. Secondly, the large spatial and temporal variability of soil water in the natural environment makes this a very difficult variable to measure on a consistent and spatially comprehensive basis.

These individual requirements have been moot before recent microwave remote sensing techniques have demonstrated a capability for measuring soil water. Both passive microwave and active microwave techniques such as synthetic aperture radar (SAR) have provided solid theoretical and experimental results that the top five cm of soil water can be measured from aircraft and space platforms under a variety of environmental conditions and through a moderate vegetation cover (Carver et al., 1989). In general, a *radar* refers to a radio-type device with a transmitter and a receiver. The transmitter illuminates a target surface and a receiver measures some property of a backscattered return signal (Ulaby et al., 1982). This paper will define the science issues and application opportunities that can be addressed by microwave remote sensing of *soil water*, describe the capabilities and limitations of the SAR systems, and illustrate one example of an application used to take this technology to an operational status.

Science Issues

Because of the ubiquitous nature of soil water in many disciplines, there are numerous potential science applications for frequent and spatially comprehensive measurements of soil water. However, most of these will fit under the umbrella of the following four science issues which are the highest priority:

- To understand the role of surface soil water in the partitioning of incoming radiant energy into latent and sensible heat fluxes at a variety of scales from mesoscale to global climate model (GCM) scale.
- To understand the relationship between the surface 5 cm of soil water observable by microwave techniques, transpiration to the atmosphere, and the total profile (1 m or more) soil water content that is accessible to plants.
- To understand how spatial and temporal patterns of soil water are related to the physical and hydrologic properties of soils.
- To understand how spatial and temporal patterns of soil water can be used to improve our ability to model runoff at a variety of scales and adapt hydrologic models to changes in ecosystems, and areas of differing climate, soil type and geology.

Application Opportunities

With the potential for measuring soil water having been demonstrated, the obvious question to ask is how might measured soil water be used by society? As in the science issues, there are four general areas in which routine measurements of soil water could have major impacts on our day-to-day lives. These are listed as follows with our best estimates of their spatial and temporal requirements:

- Improvements in medium range weather forecasting by incorporating measured soil water on a 30 km grid on a daily basis.
- Agricultural applications would include on farm irrigation scheduling, implementation of water use efficiencies, and improved crop yield forecasting for both domestic and foreign areas. The scale of interest here would be 100 m to 1 km and three to seven days.
- Water management applications require better quantification of water uses, storages and runoff to monitor existing resources and to assist decision makers in allocation of limited resources or coordination of relief efforts in times of flooding. The scales of interest here would also be 100 m to 1 km and three to seven days.
- Climate models, particularly for annual and interannual variability, need to be able to represent the land surface hydrologic processes accurately. Measured soil water can be used as a state variable and for validation of GCMs. The scales of interest here are 10 to 100 km and one to four weeks.

Capabilities and Limitations

Microwave remote sensing is of interest because of its potential for measuring soil water. The response varies with the ability of a given material to store and reflect electric energy (*dielectric constant*). For example, at long wavelengths like L-band, dry soil has a dielectric constant of about 3 while that for water is about 81. However, in a remotely sensed image the perceived effects of soil water are modulated by surface roughness, degree and direction of slope, and the vegetation cover and structure (Wang et al., 1986). Thus a corn field sloping away from the radar with rows at right angles to the flight path would appear *darker* than a similar field but one that slopes towards the radar and has the corn rows oriented in the same direction as the incident beam (Engman, 1991). The land slope effects would depend not only on the absolute value of slope but also on the resolution of the image and the angle of incoming radiation. Volume scattering from vegetation would become more dominant than the effects of surface roughness at larger (>30°) incidence angles (Mo et al., 1984).

For homogeneous soil materials, with a uniform cover, slope and structure, theoretical results have often been verified under controlled conditions (Ulaby et al., 1982). Such results increase our understanding how the microwave radiation behaves when backscattered. However, under field conditions, soil materials are seldom uniform, slopes vary, and amounts of water held by vegetation change from point to point and with time. In this study we have attempted a rather gross correction approach to minimize the impact of such variables. Because the flight path and incidence angle of incoming radiation were essentially the same on successive days of our study, the changes between images are primarily due to changes in soil water. The problem to be resolved is how to compare ground truth values of soil water measured at a point gravimetrically, with a spatially integrated remotely sensed backscatter (σ^{o}) image representing relative differences in power level and characteristics between outgoing and return electromagnetic signal in decibels (dB). We propose that one way to do this is to create a synthetic image that honors ground truth values on one hand, and corresponds closely (at least in a statistical sense) to a remotely sensed image on the other.

Objectives

The objective of this paper is to evaluate how changes in a remotely sensed image of soil water compare with changes in a synthetic image generated from ground truth values, and how the proposed corrections will affect the overall SAR image quality. The corrections applied at this time are based on *expert opinion* and readily available soil survey data using an adaptation of *multiple criteria decision-making* procedure (MCDM) described in Pereira and Duckstein (1993). We plan to develop, objective deterministic or stochatic algorithms for the same purpose. A synthetic image is assembled from measured ground truth values of soil water using a conditional simulation approach known as *simulated annealing*. The purpose is to see if the proposed procedure improves correspondence between observed changes in the remotely sensed image and measured point values on the ground. The theoretical basis relies on an extension of the Journel and Huijbregts (1978) simple *hypothesis of permanence*, which can be paraphrased as follows. Both data sets in question describe a distribution of soil water, but are expressed in different units and on a different scale. We would expect their distributions when transformed into the same units to have a similar mean but different variance depending on the size (point vs. pixel) of the measurement area, provided random error contributions due to differences in soil roughness, slope, aspect and cover can be minimized.

METHODS AND MATERIALS

Study Area

The study area described in this application example is on the Mahantango Creek watershed in Pennsylvania, USA, (Rogowski,1996). The watershed is situated in the Valley and Ridge physiographic province of the Appalachian Mountains, characterized by varying relief: upland hills, valleys, and forested mountain ridges dissected by streams. The climate in this midlatitude valley is humid with well-distributed rainfall (about 1000 mm/yr.). Elevation ranges from less then 100 m in the valleys to over 500 m on the ridge tops. The ridges, valleys and streams are oriented northeast-southwest, corresponding to the regional strike of major rock formations, while beds dip to the northwest and southeast along the centrally located anticline represented by a low ridge. The land use is predominately cropland in a rotation management system of corn, small grains, and meadow, intermingled with numerous tracts of woodland, areas of permanent pasture and orchards. Farm buildings and ponds are common, and some residential development is found along the main roads.

Soil Water

Soil water was measured gravimetrically at about 40 locations on three transects across a valley situated on a 51 ha farm (Engman and Rogowski, 1974).The primary remote sensor used in this study to measure soil water was NASA's fully polarimetric Synthetic-Aperture Radar system (AIRSAR) operated from a DC-8 aircraft. The AIRSAR is a multifrequency (P-

, L-, and C-bands), multipolarization system for observing the earth. It provides multiparameter high resolution observations in the microwave spectrum, particularly useful in assesing soil moisture, erosion and canopy structure (Carver et al., 1989,). The system relies on signal processing to achieve a narrow beam width and enhanced resolution which is independent of distance to the image area in the flight axis direction (Ulaby et al., 1982). The antenna points to the side, producing a beam that is relatively wide across the flight axis, and narrow along the flight axis. A radar pulse is transmitted from the antenna within this narrow beam. When the pulse strikes a target a portion of microwave energy scattered by the earth's surface is reflected back (*backscattered*) to be received by the radar antenna. A semicontinuous image is produced by the motion of the platform over the land surface along the flight axis. Strictly speaking the AIRSAR data are still a sample because the radar image derives from discrete, consecutive records of backscattered pulses, rather than from a recording of a continuous scene (Ulaby et al., 1982). The SAR data selected for comparison with the ground truth values were the L-band, HH-polarized readings acquired at 45° angle of incidence (measured from nadir) from overflights on July 10th when conditions were relatively dry, and from overflights on July 13th (1990), following 42 mm of rain. The orientation of soils and plant rows in this study is predominantly NE-SW. The aircraft flew parallel to the main watershed axis and the AIRSAR was looking North. Raw data were first processed by JPL and then radiometrically calibrated and georeferenced at NASA/GSFC.

Speckle and Compromise Programming

Variations in a backscattered image on a *per pixel* basis may arise from several sources. Apparent differences between pixels may be due to differences in soil water or to differences in some other pixel attribute (such as slope, aspect, surface roughness, vegetation cover or plant morphology). Each one of these attributes may affect the perceived SAR reading differently depending on local conditions. The difference may also be due to random fluctuations in a backscattered signal. This may come about as follows. In an analogy to our own way of perceiving a scene, when a radar moves past an area it can get several independent *looks*, or only a single *look* at a specific portion of the ground representing an imaginary *pixel*. In a *single look* scene the relative brightness will have an exponential distribution corresponding to a very wide range (as much as 18 dB) of random fluctuations in a backscattered signal (Ulaby et al., 1982). These random fluctuations will give rise to a so called *speckle* in the radar image. The problem is more severe with the synthetic aperture instruments and is usually addressed during processing.

To minimize the potential effects of speckle we have grouped the individual pixels into three by three 12 m pixel blocks, averaging the nine values. To minimize the effects of attributes other than soil water we have implemented a multiple criteria decision making approach (MCDM) based on the *compromise programming* (CP) detailed in Pereira and Duckstein (1993) and Zeleny (1982). The *compromise* aspect is usually expressed in a form of *distance metric* (d_p) which describes the degree of sameness between some observed and ideal value. The algorithm used to compute the distance metric is:

$$d_p = \left[\sum_i^I \beta_i^p \left\{ 1.00 - wt_i^k \right\}^p \right]^{1/p} \quad (1)$$

where wt_i^k denotes a weight assigned to a given category of an attribute, *i* refers to the four attributes: slope, aspect, land use and surface roughness; *k* denotes individual categories; β_i describes the relative attribute importance; and d_p is the distance metric correction factor for p=1, 2, and 10 $(p \approx \infty)$. The approach has been utilized in many different and seemingly unrelated fields. Pereira and Duckstein (1993) used the concepts of MCDM and d_p in a land suitability evaluation of a squirrel habitat, Bogardi et al. (1985) employed it in a well network design, while Burrough (1986) discussed it in connection with map classification and cluster analysis of soil properties. Here it is used as a correction factor to modify gross effects due to different categories of land slope, aspect, soil roughness and vegetation. Because our soils tend to be very stony, the degree of stoniness is used as a surrogate for surface roughness. In the context of a GIS each of these attributes is considered as a separate, spatially varying overlay (Figure 1). Table 1 illustrates a possible schedule of weights (wt_i), and lists the attribute importance rating (β_i). The correction factor aspect is implemented by *subtracting* each category weight (wt_i) from 1.00 in Eq. 1 and setting the ideal category weight (wt_i) to 0.00. The category weights (wt_i), were chosen by individuals familiar with workings of radar imagery and properties of the study area. Because some attributes may more strongly influence the backscattering coefficient (σ^o) than others, we introduced, following Pereira and Duckstein (1993), a weighting coefficient β_i which indicates a relative importance of a given attribute, with the sum of β_i values equal to 1.00. The computed distance metric (d_p) for p =1, 2, and 10, suggests the extent of potential change (reduction coefficient) in the observed SAR values.

The d_p-algorithm was processed on a pixel-by pixel basis by the OVERLAY, SCALAR and TRANSFOR modules of the GIS IDRISI (Eastman 1992) [the mention of trade names does not constitute an endorsement of the product by the US Department of Agriculture over other products not mentioned]. Individual blocks of 3×3 original pixels were then corrected using an appropriate value of d_p for p=1, 2 and 10 (∞). Figure 2 shows distributions of the correction factor d_p for different values of p, as quartiles (Q) of the distribution while Table 2 lists the corresponding reduction coefficients and statistics. When p=1, d_p is the so called *city block* distance metric which assumes total compensation among criteria, with all contributing equally (Zeleny, 1982). Thus large weights (wt_i) of some attributes in Table 1 could be compensated by the small weights of others. The p = 1 correction will assign a larger area to higher values of d_p, and amount to a lesser overall correction of the SAR data. In contrast, when p = ∞ (i.e. any value of p $\geq$ 10) there is no compensation among criteria. Reduced d_p values result due to the overriding influence of categories which carry the largest weight. This is illustrated in Figure 2. For example, southern half of the study area is composed of mainly of Berks and Hartleton soils. Their large weights (Table 1) are compensated for by the northern (N) exposure with the result that more area is assigned to a higher value of d_p (blue). This leads to lesser overall correction (0.81 reduction coeficient) in Table 2. The opposite is true when p = ∞. The area in Figure 2 that was *blue* for p = 1 is now yellow, and the overall reduction in Table 2 is now equal to 0.28. Finally, when p = 2 we have a so called *shortest distance* or *linear* d_p with only a partial compensation, and an effective extent of image correction falling somewhere in-between the correction for p = 1 and p = ∞.

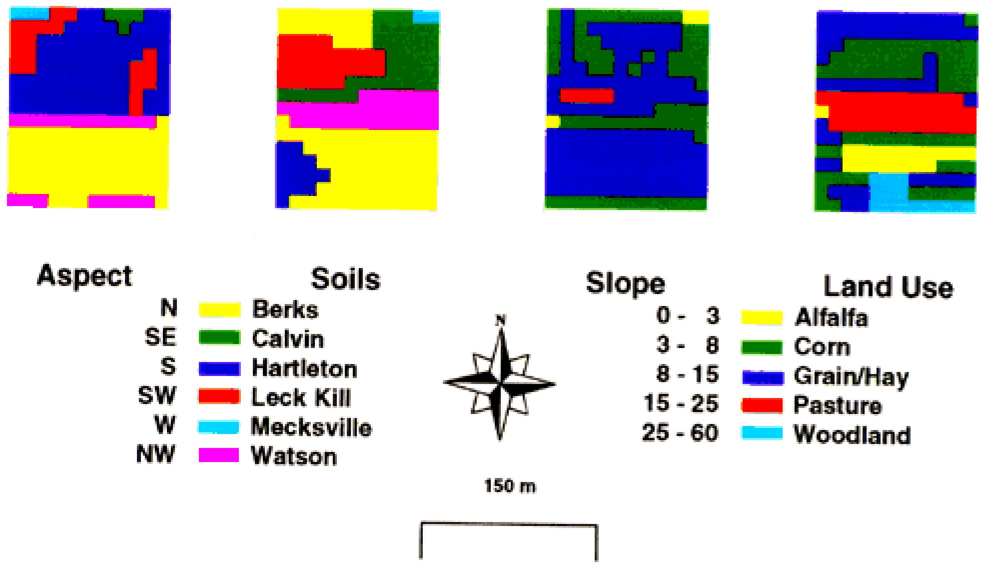


Figure 1. Spatial distribution of four terrain attributes: slope, aspect, soils and land use, over the study area.

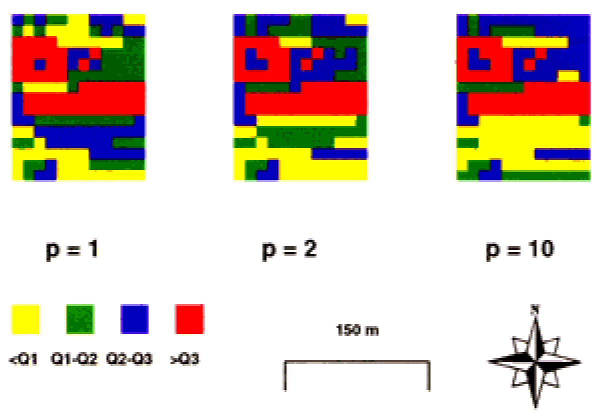


Figure 2. Spatial distribution of the four quartiles (Q) listed in Table 2 of the correction factor represented as a distance metric (d_p) reduction coefficient for p=1, p=2, and p=(∞) (any value greater than or equal to 10).

Table1. Schedule of weights (wt_i) for each category (k) of four attributes (i), used to computed a correction factor as a distance metric (d_p); relative attribute importance is given by β_i.

Slope		**Aspect (1)**		**Land Use**		**Soil**	
0-3%	0.00	None	0.00	Stubble (2)	0.00	Leck Kill	0.00/<5
3-8%	0.06	N	-0.20	Pasture	0.06	Watson (3)	0.10/>5
8-15%	0.14	NE/NW	-0.10	Alfalfa	0.10	Calvin	0.33/>8

15-25%	0.23	W/E	0.10	Corn	0.17	Berks	0.50/>10
25-60%	0.26	SE/SW	0.30	Orchard	0.28	Hartletn	0.66/>15
>60%	0.29	S	0.50	Woods	0.33	Mecksvl	0.80/>20
β_{slope}	0.28	β_{aspect}	0.13	$\beta_{land\ use}$	0.26	$\beta_{roughness}$	0.33

1. Compass orientation of slopel
2. Small grain stubble (barley, wheat, oats) and mowed hay.
3. Soil series used as surrogate for roughness bassed on stone content (wt/ %fragments >75 mm).

Table 2. Statistics (1) of the distance metric (2) correction factor d_p, for p=1, 2, and 10.

	$d_{p=1}$	$d_{p=2}$	$d_{p=10}$		$d_{p=1}$	$d_{p=2}$	$d_{p=10}$
	reduction coefficient						
Q_1	0.76	0.40	0.26	Skew	0.66	0.61	0.59
Q_2	0.79	0.41	0.27	Kurt	3.06	1.99	1.68
Q_3	0.87	0.47	0.33				
Mean	0.81	0.42	0.28				
SD	±0.07	±0.04	±0.03				

1. The 25th (Q_1), 50th (Q_2), and 75th (Q_3) quartiles, Mean and Standard Deviation (SD); coefficients of Skewness (Skew) and Kurtosis (Kurt)
2. Correction factor as distance metric $d_{p=1,2,\text{and }10}$ in % reduction.

Simulated Annealing

Accurate calibration of the microwave backscattering coefficient (σ^{o}) with measured soil water data is difficult, largely because of uncertainty in other terrain properties and differences in a sampling scale. Soil water is usually sampled at a point in relatively few locations, while calibration of the AIRSAR image calls for a spatial distribution of soil water over a much larger area. We have attempted to develop a preliminary correction of the AIRSAR image for gross effects of such attributes as slope, aspect, land use and surface roughness as well as possible effects of radar speckle. This was illustrated and discussed above in the context of the MCDM and distance metric $d_{p=1,\ p=2,\ p=\infty}$. Assuming the above corrections can account for some of the differences, observed in the field, we propose to use a conditional simulation approach known as *simulated annealing* (SAS) (Deutsch, 1992), to produce a spatial distribution of soil water based on measured values which is similar to an AIRSAR distribution (at least in a statistical sense).

Simulated annealing (Deutsch and Journell,1992) is a stochastic optimization technique analogous to the physical process of annealing. It can be used for modifying a stochastic image generated from conditioning data to meet a pre-specified objective function. The objective function (*O*) used in our simulated annealing program is,

$$O = \sum_h \frac{\left[\gamma * (h) - \gamma(h)\right]^2}{\gamma(h)^2} \quad (2)$$

where $\gamma * h$ is the semivariogram of the synthetic image and the $\gamma(h)$ is the semivariogram of the AIRSAR image. The simulation can reproduce two-point statistics of the target distribution and combine them with multivariate spatial statistics obtained from training images, while honoring the original field derived soil water data.

In a physical process of annealing a solid is heated sufficiently high to melt. If it is then allowed to cool very slowly, it crystalizes into a *ground state* with a perfect lattice and minimum free energy. The ground state is obtained only if the initial temperature is sufficiently high and if the cooling is sufficiently slow (Aarts and Korst, 1990). The physical process can also be modelled on a computer using a general approach proposed by Metropolis et al. (1953) and based on a following criterion,

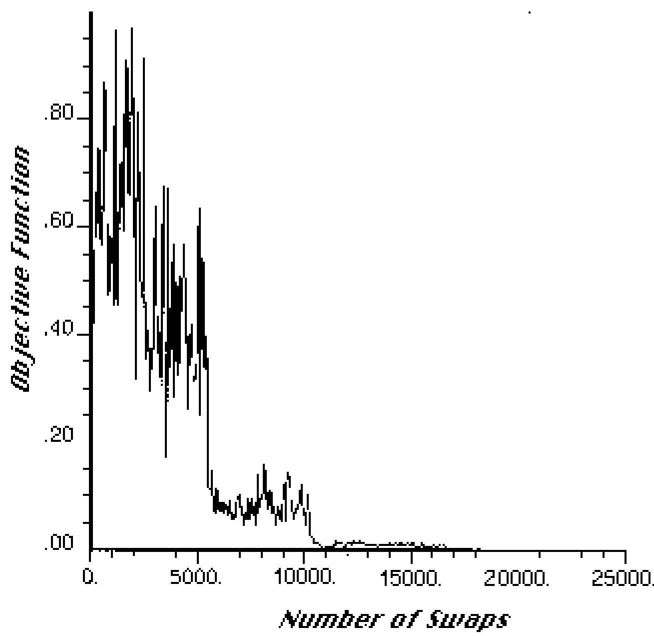
$$\exp\left(\frac{E_i - E_j}{k_b T}\right) \quad (3)$$

where E_i is the energy level of a current state, E_j is the energy level of the next state following a small lattice perturbation, k_B is a Boltzman constant and T refers to an ambient temperature. Any change in a configuration of a lattice will bring about a change in energy $(E_i - E_j)$. If E_j is less than E_i the change is always accepted, if not it may still be accepted with a probability given by Eq. 3 which is known as *Metropolis criterion* and is embedded in a *Metropolis algorithm*. The simulated annealing uses the Metropolis algorithm to obtain a sequence of solutions to a combinatorial optimization problem (Aarts and Korst, 1990). The analogy assumes that *solutions* are equivalent to physical *states*, and *costs* reflect corresponding physical *energy levels*.

In applying simulated annealing to AIRSAR problem we will assume that simulated scenes are equivalent to the *solutions* or *states* mentioned above; perturbations to a *lattice* correspond to changes in pixel backscatter while changes in the *energy level* (Eq. 3) are reflected by changes in the objective function *O* (Eq. 2). To implement the SAS we proceed as follows. First we relocate the conditioning data (consisting of soil water content measurements within the study area) to the grid nodes of 3×3 pixel blocks. Because we will need the AIRSAR distribution as a pdf source of data we have to convert the conditioning Θ_s data into the same units as the AIRSAR values. This can be done using for example Wang et al. (1986) equation. We then assign the remaining pixel values at random by drawing from a pre-specified probability distribution function (pdf) which in our case corresponds to the original AIRSAR image. The synthetic image is then modified by interchanging (*swaping*) grid nodes that do not involve the conditioning data and updating the image.

The interchange is always accepted if the value of objective function is lowered. It may also be accepted with a Boltzman distribution involving Eq.3 (Aarts and Korst, 1990) for some less favorable changes when *O* increases, if the annealing temperature parameter is high enough. This is illustrated in Figure 3 where the value of objective function is plotted against the number of *swaps*. The procedure (Deutsch, 1992; Deutsch and Journell, 1992) calls for an *annealing schedule* to specify parameters which describe how a *temperature* (parameter T in Eq. 3) is to be lowered and when to terminate the simulation. The choice and coding of objective function *O* offers a number of possibilities. To illustrate, we use the semivariogram γ_h formulation available in GSLIB (Deutsch and Journell, 1992). The objective is for the semivariogram of a synthetic image in Eq. 3 to match that of the AIRSAR image. When the simulation is complete, we back transform the simulated values and post-process them with a uniform probability field (Srivatava, 1992) based on AIRSAR data, by assuming that their local magnitudes correspond to the conditional probability frequencies.

July 10 Objective Function versus Swaps



- Number of data: 498
- X Variable
    - mean: 12291.990
    - std. dev. 7118.731
- Y Variable
    - mean: .141
    - std. dev. .242
- correlation: -.743
- rank correlation: -.964

Figure 3. The objective function *O* plotted against the number of swaps; sudden drops indicate times when temperature is lowered.

RESULTS AND DISCISSION

Illustrative example

The study area we use to illustrate the procedure outlined above, represents a 12×15 arrangement of 180 blocks, with each block based on the nine AIRSAR pixels. For most blocks (~160), the value of a block was an average of nine pixels. However, for blocks used as conditioning data the AIRSAR value most closely resembling a value computed from the Wang at al. (1986) equation ($\sigma^{o}_{estimate}$) was chosen as block value. This was done because the conditioning data had to be of a similar magnitude and in the same units as the backscatter values. Our AIRSAR values were found to lie between those in Wang et al., (1986) and those in Ulaby et al., (1982, p.861) equations. Empirical equations, such as Wang et al., (1986) regression equation, have been used to relate a backscatter coefficient σ^{o} in dB to gravimetric soil water content in percent,

$$\sigma^{0} = -25.7 + 0.302\Theta_{g} \quad (4)$$

The Ulaby et al., (1982) regression equation may be represented in similar terms, assuming average gravimetric field capacity of soils in the study area to be 0.2075,

$$\sigma^{0} = -15.96 + \left(0.148\ \Theta_{g} / 0.2075\right) \quad (5)$$

Data for Eq. 5 were acquired from a truck mounted radar spectrometer and are for bare soil with a variety of surface roughness and texture conditions. Data for Eq. 4 highlight differences in the land use and were observed from a space platform. Our AIRSAR values correspond to backscatter in a broken and sloping terrain, with different land uses typical of the NE USA. Because large stone fragments are more abundant in some areas than in others, surface roughness and soil water content vary appreciably from point to point. Field measured values of Θ_{g}, when substituted in Eq. 4, provided estimates of σ^{o} ($\sigma^{o}_{estimate}$) used in selection of the conditioning values from an array of the nine surrounding AIRSAR data.

Table 3. Univariate Statistics (1) for the SAS and SAR distributions.

SAS (2)			SAR (3)		SAR			SAS	
Date	10th	13th	10th	13th	Date	10th	13th	10th	13th
Q_1	-21	-18	-20	-17	Skew	0.82	0.61	0.82	0.55
Q_2	-18	-16	-18	-16	Kurt	0.83	0.55	0.84	0.25
Q_3	-16	-14	-15	-13	W:N	0.95	0.97	0.95	0.97
Mean	-17.9	-15.8	-17.7	-15.5					
SD	±5.3	±3.1	±5.7	±3.4					

1. The 25th(Q_1), 50th (Q_2), and 75th (Q_3) quartiles, Mean and Standard Deviation (SD) in dB; also coefficients of Skewness (Skew) and Kurtosis (Kurt), and the Wilkes-Shapiro Statistic W: Normal (W:N).
2. Simulated annealing simulation (SAS) of soil water, transformed into units of dB using Wang et al., (1986) Equation 5a and the average value of FCg = 0.2075.
3. The original SAR data.

Image Correction

Results in Table 3 compare AIRSAR values (SAR) with those derived from *simulated annealing* (SAS). The corresponding SAR and synthetic SAS images are shown in Figure 4. The synthetic SAS images were conditioned by 22 point values measured as transects in the field on July 10th and by 26 point values measured as transects on July 13th following 42 mm of rain. The table lists the respective univariate statistics of simulated data and the corresponding values of AIRSAR backscatter coefficient (σ^o) in dB. Figure 4 illustrates a corresponding spatial distribution of the observed and simulated values and the changes which occurred between the July 10th and 13th following 42 mm of rain. Results suggest that a SAS generated image of properly transformed data will resemble the SAR image statistically and structurally but not necessarily visually. This is because the objective of stochastic simulation is not to reproduce exactly the original image, but rather to extract its important points with limited statistics (Deutsch, 1992). The starting distribution is drawn at random from the AIRSAR pdf and corresponds closely to the image values only at the conditioning data locations. Significantly the simulated distributions conditioned by point Θ_g data do have a similar mean and standard deviation (variance) as the original AIRSAR image.

To obtain soil water content corresponding to the AIRSAR values in Figure 4, the SAS images were first converted into gravimetric soil water content (Θ_g) using Wang et al., (1986) equation (Eq. 4). The values of (g obtained for each pixel are plotted against corresponding AIRSAR values in Figure 5a. The data in Figure 5a show no correlation between the backscatter coefficient σ^o from remotely sensed SAR images on the 10th and 13th of July, and soil water content. All that can be said is that they fall between those in Wang et al., (1986) and Ulaby et al., (1982) equation. However post processing Θ_g values with a probability field (P_{field}) leads to a better correlated (0.42) distribution (Figure 5b) somewhat larger than Wang et al., (1986) values. The post processing was accomplished by generating a uniform probability field (Srivatava, 1992) between 0 and 1 from a uniform scores transformation of the AIRSAR image. The resultant image was further adjusted in Figures 5c, 5d, and 5e by imposing a correction for dp with p=1, 2 and ∞ on data in Figure 5b. The net result was an overall shift towards the Ulaby et al. (1982) equation.

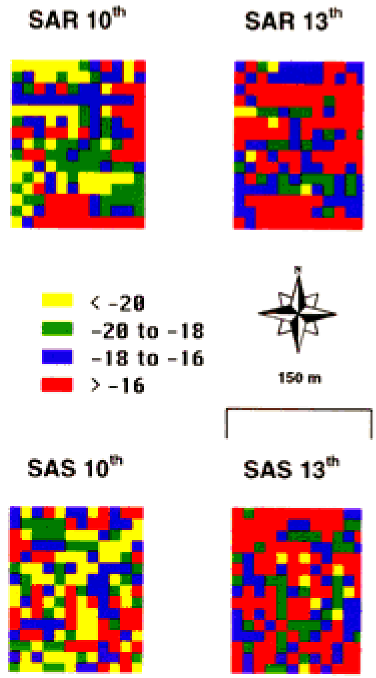


Figure 4. Spatial distribution of the observed (SAR) and simulated (SAS) backscatter coefficient (dB) on the 10th and 13th of July,1990.

MCDM Compromise Programming

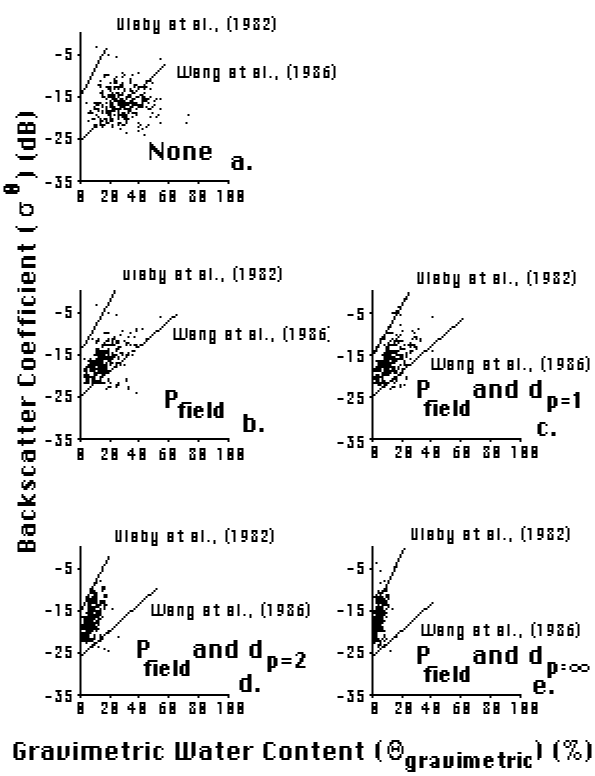


Figure 5. Distribution of AIRSAR values as a function of gravimetric water content (Θ_g) obtained from simulated annealing (a); post-processed with a Probability Field (Srivastava, 1992) based on AIRSAR data (b); and adjusted for d_p, with p=1 (c); p=2 (d); and p=∞ (greater than or equal to 10) (e).

There are currently a number of regression equations available (Ulaby et al.,1982) which relate a backscatter coefficient (σ^o) in dB to soil water content. While not identical to that proposed by Wang et al., (1986) they generally differ by a *factor*. We propose that a correction factors can be estimated using an approach, described by us in this paper. We note that our correction factors are based on available Soil Survey data in order to illustrate our approach. In reality there is no guarantee that any attributes such as *soil*, *slope*, *aspect* and *land use* at a point are as stipulated by the survey. Thus, it may be advisable to sample actual distributions of these attributes in the field, and base all simulations on field derived values. Terrain and location attributes at a 3×3 pixel scale may be considered as random variables. The SAS simulation of such attribute overlays would tend to generate a random correction factor, which would be more in keeping with field observed values.

CONCLUSIONS

Most research to date has indicated that the use of single frequency, single polarization SAR data may have limited use as soil water sensor because of inability to decouple soil roughness and land use effects from those of soil water. The approach taken attempts to remedy this situation. The procedure consists of generating statistically equivalent images of soil water content based on relatively few measured values. These are subsequently postprocessed with a probability field (Srivastava, 1992) derived from AIRSAR data. The resultant distribution may be further adjusted for possible effects of slope, aspect, land use and surface roughness using a multicriteria decision making (MCDM) approach known as compromise programming (CP). The procedure leads to a fair correspondence between field measured and remotely sensed data and renders soil water changes under natural conditions more quantifiable. Current approach relies on expert knowledge to choose approriate weights for the MCDM correction. However, since the sensitivity of radar backscatter to the terrain attributes is well documented in literature, the choice of weights may eventually be replaced by more objective deterministic, or stochastic algorithms now under development.

ACKNOWLEDGMENTS

We would like to acknowledge Don Simmons for data manipulation and processing , Ann Hsu for AIRSAR data calibration and georeferencing, and Eric Warner for obtaining ground truth point coordinates in the field and making them available to us.

REFERENCES

Aarts, E. and Korst, J. (1990) *Simulated Annealing and Boltzman Machines.* John Wiley & Sons, New York, NY 2nd Edition. 272p.

Bogardi, I., Bardossy, A., and Duckstein, L. (1985) Multicriterion network design using geostatistics. *Water Resources Research* 21(2):199-208.

Carver, K. et al. (1989) *SAR Earth Observing System: Instrument Panel Report.* Vol.IIf & Appendix, NASA TM, 233p.

Burrough, P.A. (1986) *Principles of Geographical Information Systems for Land Resources Assessment.* Monographs on Soil and Resources Survey NO 12. Clarendon Press, Oxford. 194 p.

Deutsch, C.V. (1992) *Annealing Techniques Applied to Reservoir Modeling and the Integration of Geological and Engineering (Well Test) Data.* Ph.D. thesis, Stanford University, Stanford, CA. 306 p.

Deutsch, C.V. and Journel, A.G. (1992) GSLIB *Geostatistical Software Library and User's Guide* Oxford University Press, New York. 340 p.

Eastman, J.R. (1992) *IDRISI a grid based geographic analysis system.* Version 4.1. Clark University Graduate School of Geography, Worcester, MA 01610. 178p.

Engman, E.T. and Rogowski, A.S.(1974) A partial area model for storm flow synthesis. *Water Resour. Res.* 10(3):464-472.

Engman, E.T. (1991) MACHYDRO-90; the microwave aircraft experiment for hydrology. *Int.Geosci.Remote Sensing Symposium* Vol.II p.749-751. Espoo, Finland. IEEE Cat.No.91CH2971-0

Journel, A.G. and Huijbregts, CH.J. (1978) *Mining Geostatistics.* Academic Press, New York, NY. 600p.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21: 1087-1092.

Mo, T., Schmugge, T.J. and Jackson, T.J. (1984) Calculations of radar backscattering coefficient of vegetation covered soils. *Remote Sensing of Environment* 15:119-133.

Pereira, J.M.C. and Duckstein, L. (1993) A multiple criteria decision-making approach to GIS-based land suitability evaluation. *Int.J.Geographic Information Systems* 7(5):407-424.

Rogowski, A.S.,1996. Quantifying soil variability in GIS applications. II.Spatial distribution of soil properties. *Int.J.Geographic Information Systems* (In Press)

Srivastava, R.M.(1992) Reservoir characterization with probability field simulation. *The 67thAnnual Technical Conference and Exhibition of the Society of Petroleum Engineers. SPE 24753:927-938.*

Ulaby, F.T., Moore, R.K. and Fung, A.K. (1982) *Microwave Remote Sensing: Active and Passive:* Volume II, Addison-Wesley, 607pp.

Wang, J.R., Engman, E.T., Shiue, J.C., Rusek, M.and Steinmeier, C.(1986) The SIR-B observations of microwave backscatter dependence on soil moisture, surface roughness, and vegetation covers. *IEEE Trans.Geosci.Remote Sens.*, GE-24: 510-516.

Zeleny, M. (1982) *Multiple Criteria Decision Making* McGraw-Hill Book Co., New York. 563 p.

Andrew S. Rogowski USDA-ARS-PSWMRL, University Park, PA 16802-3702, USA.
TEL: 814-863-8758, FAX: 814-863-0935, EMAIL: asr2@psu.edu

Edwin T. Engman Hydrological Sci. Branch, Laboratory for Hydrospheric Processes,NASA/GSFC, Greenbelt, MD 20771 USA.
TEL: 301-286-5480 FAX:301-286-8522 EMAIL: tengman@neptune.tsic.nasa.gov

Towards a Methodology for Selecting a "Characteristic" Sample from an Existing Database: An Evolutionary Approach

Jarvis, C.H.*, Stuart, N.*, Kelsey, J., Baker, R.H.A.

ABSTRACT

The performance of an environmental model depends highly on the 'representativeness' of the sample data with which it has been developed in relation to the overall data set, together with the adequacy of the sample with regard to the modelling technique itself. While advances have been made towards achieving both goals individually, in few cases are these combined. Additionally, the situation with regard to pre-existing data at fixed points remains 'ad hoc'. Given that increasing volumes of pre-sampled point data now exist in electronic form, these omissions are an issue of currency within the wide range of topics embraced by the term 'environmental modelling'. That this increase in data richness has frequently been matched by the rising cost of data means the ability to select appropriate data is paramount not only in terms model quality but also in the economic feasibility of the overall project itself. A new sampling methodology is presented to deal with these issues, framed within the context of the choice of meteorological station data for the development of interpolated weather surfaces but of wider applicability.

Fundemental to the new strategy is the representation of sampling requirements by means of a multi-criteria function. Criteria are developed to deal with each individual variable, for example a desired height function/range or the maximisation of record length. Data selection may also be tailored to the model types for which the data is to be used, for example by ensuring that an adequate number of nearby points are selected for interpolation by kriging. This combination of possibly contradictory criteria are then optimised together using a genetic algorithm in conjunction with the geographical information system in which the data are stored. While the solution of multi-criteria problems can be achieved using a variety of more traditional techniques, the GA approach was chosen both for its ability to manage large volumes of data and for its management of compromise. As a working progress report, discussion focuses on the reasons for the choice of an evolutionary approach and the representation of the sampling problem within this framework.

INTRODUCTION

However sophisticated a model or procedure, the results will only be as good as the data/knowledge underpinning its conception. This statement in itself is not new or radical and a number of sampling methodologies are very familiar in a wide range of disciplines. Such techniques include for example the random, stratified or systematic families of sampling. More recently these familiar methods have been augmented by those such as gradsect sampling (Austin & Heyligers, 1989) which deliberately sample across steep gradients of environmental change. Such moves highlight an increased awareness of the need to maximise the possible environmental range within a sample. Additionally, work by biologists Margules & Stein (1989) stressing the limitation of sampling within one dimension only has been applied within a geographical framework (Lees 1994, Aspinall & Lees 1994). The latter argue that not only should sampling be carried out with respect to full range of environmental criteria, but that each environmental space should be sampled individually. Further dimensions to the sampling problem that are difficult to incorporate within the context of traditional sampling methodologies include the purpose for which the data is to be used (e.g. Pettitt & McBratney, 1993), together with strategies which formally take into account the cost of sampling within the overall sampling objective (e.g. Bras and Rodriguez-Iturbe, 1976). What ma y be summarised from the current literature therefore are the concepts of representative data, the need to look at multidimensionality and the possibility of tailoring our sample according to future analytical requirements.

In the case of choosing the most characteristic pre-located site data from a wider set, as distinct from a free choice of sample sites over a particular landscape, the means by which all of the principles discussed above may be applied within the sampling process is somewhat ad hoc. Such a situation is however frequently to be found with the increasing volumes of pre-sampled data available in electronic form, the task of choosing relevant data becoming commensurably awesome. Even when a restricted and relevant data exists, its full cost may deem a project economically not viable and therefore a means of partial selection critical. Applying the definition of Eastman (1995) which classifies problems with a single objective (making an 'appropriate' sample choice) subject to a number of possibly conflicting criteria (representative data, multidimensionality, analytical requirements, cost) as multi-criteria evaluations, techniques for solving such problems within the broader literature are used as a base for further methodological discussion. It should be noted that some ambiguity in the use of such terminology arises even within a GIS environment, Jankowski (1995) referring to both multi-criteria and multi-objective techniques as 'multiple criteria decision making methods'.

A wide variety of traditional optimisation and search techniques exist which have been drawn upon in the development of 'multiple criteria decision making methods' (Table 1). As Schwefel (op cit, p165) notes, 'the question of which is the best strategy is itself a kind of optimisation problem' ! Just as multi-objective analysis has tended to be viewed as a 'natural extension of mathematical programming' (Jankowski 1995), so have the overlay and hierarchical methods of multi-criteria analysis found favour within the context of a geographical information systems (e.g. Carver 1991, Eastman et al 1995).

Gradient analysis Enumerative

Mixed Integer Programming

Dynamic Programming

Overlay methods

Hierarchical methods

Random

Random walks

Random search & save

Table 1, CONVENTIONAL SEARCH AND OPTIMISATION METHODS (After Goldberg, 1989, p3)

While a straightforward solution to the multi-criteria sampling problem was sought, a number of disadvantages in using these conventional methods were identified. These are summarised in Table 2 below.

Gradient analysis:

Search spaces are often calculus unfriendly i.e. non-differentiable and are difficult to define.

Local in scope (e.g. calculus (or gradient) methods) and may therefore converge prematurely at a non global solution

Not designed with multiple solutions in mind, which occur with competing objectives, and are therefore devalued as a decision support tool

Often highly tailored to solve a particular problem, so both may not be easily adapted and fail to enhance the GIS in other potential application areas

Not adaptable to situations with multiple solutions without many runs (Fonseca & Fleming, 1995)

Enumerative methods:

Solutions problematic to understand with increasing number of criteria (e.g. overlay analysis, Janssen & Rietveld 1990)

Loss of information using thresholds with continuous data & problems in defining thresholds in overlay analysis (Janssen & Rietveld op cit.)

Unable to compromise between conflicting criteria (e.g. hierarchical methods)

Not guaranteed to find a solution (e.g. Overlay/weighted linear combination methods or hierarchical ranking methods may render the entire search space 'unsuitable')

May not address combinatorial side of a problem (e.g. overlay/hierarchical methods ignore geographical relationship between possible sites, important in sampling context)

Unable to cope with large volumes of data (incl. dynamic programming, Schwefel 1995 p12)

Not necessarily adaptable to situations with multiple solutions in mind (e.g. simplex, Fonseca & Fleming 1995)

Random search:

Unable to cope with large volumes of data (choosing 200 sites from a total of 985 has 985!/((985-200)!*200!) possible combinations)

Table 2, RESTRICTIONS OF CONVENTIONAL ALGORITHMS

In addition to conventional search and optimisation algorithms, a newer class of methods known as evolutionary algorithms (incl. genetic algorithms) has also been used in multi-criteria optimisation (Goldberg, op cit, p197). Because of the major obstacles emerging should traditional methods be used in the case of the sampling problem (search space size, the objective as a collective of suitable sites, and conflicting criteria selection) these more recent techniques are evaluated for their potential use in the sampling application. The results of this analysis are shown in Table 3, from which the decision to develop a genetic algorithm approach to the extraction of a characteristic sub-sample of fixed data points from an established database was taken.

Work from a large number of points simultaneously, so climbing many 'local' peaks in parallel and therefore improving upon simulated annealing in addition to traditional hill climbing methods (Michalewicz, 1992, p29)

No auxiliary information in the form of derivatives is required, although the prior tabulation of distances between sites is needed for efficiency

Makes use of guided but randomised search to give wider coverage of data space than that provided by deterministic methods (Schwefel op cit p109)

Use of codings of parameter set, rather than parameters themselves, means that GAs are largely unconstrained by previously mentioned limitations such as continuity, derivative existence etc. (Goldberg, op cit, p7)

General, rather than specific, technique allows code and methodology to be used for a wide variety of tasks when coupled with a GIS subject to alteration of codings/objective functions (travelling salesman e.g. Goldberg op cit, p170; groundwater pollu tion containment, Ritzel et al 1994; facilities siting, Pereira et al 1993)

But, theoretical study of multi-criterion evolutionary algorithms is lacking (Michalewicz, op cit, p9)

Table 3, ADVANTAGES OF EVOLUTIONARY COMPUTING APPROACH

METHODOLOGY

In developing a new sampling methodology, the first question to be asked is 'what is required of the sample data?'. As established within the literature review, the two main goals should be the 'representativeness' of the data with regard to the overall problem space, and the tailoring of data to suit the requirements of further analytical techniques. Specific criteria relevant to the search may then be introduced within the framework of more general evolutionary code. The local context of the sampling problem therefore requires close analysis. In this case the use of the sample is in the interpolation of point type meteorological data for England & Wales. The problem task is to choose a total of 200 sites from a possible 985, the number of sites restricted for economic reasons. The available meta-data for each site, derived data and its source are detailed below (Table 4) and in the first instance exclude partial or subjective data such as quality statements.

MET. OFFICE DERIVED	GIS
Location	Location
Start of recording (year)	Record length
End of recording (year)	Age of information
	Currency of data

Table 4, SOURCES OF META DATA REGARDING EACH SITE

Derivations have been developed within the GIS GRASS or by spreadsheet, and move the initial data closer to a form useful in meeting the sampling criteria which are outlined below

A. Goal 1: 'Representativeness' of sample data relative to full data set

Adequate coverage of sites throughout country

Distribution of aspects representative of those of UK as a whole

Distribution of heights representative of total range within UK

B. Goal 2: Sampling requirements related to interpolation tasks

Nearby sites required for successful production of variogram

Long records required for improved temporal predictions

Currency of data important for management of future infestations

Where data recording has stopped but site important for other characteristics, time since recording discontinued

Table 5, DEVELOPMENT OF SAMPLING CRITERIA

Straightforward formulation of the criteria into 'fitness' functions does not however necessarily imply ease of optimisation, as hinted at by the ease in which more traditional approaches may be discounted in this problem case. A number of methods o f multi-criterion optimisation using evolutionary techniques exist, and this subject in itself forms an area of active research. Indeed, Fonseca & Fleming (1995) suggest that for the related multi-optimisation problem that it is time that an experimental approach be taken to real-world problems. Choice of method incorporates the two main issues involved within the evolutionary approach, those of fitness assignment (or suitability of the site) and search strategies (the probabilistic search through disc rete space). In the main, attention has been given to the former (e.g. Jankowski 1995) and thus this is where the main alternatives lie. Given the apparent lack of comparative material and a desire to maintain simplicity where possible, the most straightforward of approaches is chosen firstly: the use of the popular aggregation (weighted sum) approach to fitness familiar from other techniques together with standard (Michalewicz, 1992) search operations.

The basic implementation of the evolutionary approach is summarised below (After Davis, 1991):

- Initialise a population of chromosomes (Where each chromosome is composed of 200 randomly selected sites (genes), selected from the possible 985 sites)

  e.g. site1, site3, site16, site89, site59, site895 ... gene 2 = site 3, gene 5 = site 59

  A floating point rather than the more conventional binary representation is used to prevent prohibitively long strings. This also has other advantages: it is intuitively closer to the problem space, facilitates the design of problem specific operators, and is faster and more consistent (Michalewicz, op cit, Chap. 5). This means that the approach used is not strictly that of a 'pure' (Goldberg 1989) genetic algorithm, hence the use of the broader term evolutionary algorithm..

- Evaluate each chromosome in the population according to its fitness (Fitness assignment) (Where fitness equals the weighted sum of each of the seven (normalised) individual criteria fitnesses, akin to approaches developed for use with conventional optimisers)

  eg. Height fitness for chromosome 1 within the population (population = 200)

  f(height)= (max. height - min height)/max. range possible within all 985 sites

  Create new chromosomes by mating current chromosomes, using procedures of crossover and mutation (Search strategy) and where the probability of breeding relates to fitness.

  eg. simple, single point crossover (crossover point = |)

  Initial two strings: s1, s4, s6 | s768, s500, s345 ... , s98, s687 | s27, s879, s56 ...

  Result: s98, s687, s768, s500, s345 ... s1, s4, s6 s27, s879, s56 ...

  eg. mutation.

  A value selected for mutation is replaced by a randomly generated number between that variable's lower and upper bounds

  In both cases, the possibility exists to create illegal combinations, that is strings in which a particular site appears more than once. For the promotion of diversity, no repair algorithm was applied to the crossover operation on the basis that the fitness function rewards a spread of points. The basic algorithm was however altered to prevent any illegal mutations from occurring, and additionally f(nearness) was amended to avoid rewarding duplicate sites.

  A further issue arises when the contribution of one gene to fitness depends on the value of another gene (Epistasis). This can lure the evolutionary algorithm into suboptimal convergence, as explained by Michalewicz (op cit, p52). Because this situation arises in the case of the height fitness for example, an inversion operator was incorporated within the methodology which aims to group the maximum and minimum height genes together within a string to form an important building block.

e.g. inversion

Initial string: s1, s4, s78, s627, s3, s783, s876 ... , max. height = s4, min height = s783

After inversion: s1, s4, s783, s78, s627, s3, s876 ..

- Delete members of the population to make room for new chromosomes
- Evaluate new chromosomes and insert them into the population
- If the set number of generations is up, terminate and return the best chromosome
- Repeat experiment five times and average results

RESULTS AND DISCUSSION

In order for the evolutionary approach to sampling to be evaluated , its advantages and improvements compared to older strategies must be displayed. A major problem facing the study in this regard is that the performance of the evolutionary algorithm is a vector. Because of conflicting criteria, such as the requirements for sites to be spread throughout England and Wales together with the reward of nearby sites in order to be able produce an adequate variogram, a number of non dominated solutions are likely to exist owing to compromise. How therefore can a set of runs be evaluated? Perhaps only by the relative performance of the sample data in producing an interpolated data surface. However, because this experiment is set within the context of a real requirement for meteorological data, only meta-data is as yet available.

Instead, average initial results of the overall fitness of the solution were compared to that from a simple deterministic methodology in which sites were 'weeded' in a sequential fashion according to each criterion in turn, ranked according to perceived importance. While using this comparative technique selections could be made relatively straightforwardly according to second derivative of slope, age and continuance however, problems arose in incorporating the full range of environmental parameters such as height and in balancing a good spread of sites with adequate nearest neighbours in geographically varied areas. The overall fitness value achieved by the ad hoc technique compared well with the 'best' value of the initial random populations of the evolutionary algorithm but unsurprisingly was not well balanced within the scores for individual criteria.

One of the obstacles to the wider application of evolutionary algorithms is the need to experiment using a number of parameters such as the mutation rate, the probability of crossover between populations, population size used and the form of the operators themselves. The success of the evolutionary procedure may be shown by reference to the average convergence curve of fitness per generation over several runs using identical parameters from different random start points and additionally the use of different strategies from similar start points. Work in progress currently shows a steady increase in the average fitness of the total population per generation, with slower gains in 'best' fitness. This steady increase in average population value indicates a performance better than purely random and that the methodology holds promise. Nevertheless, it is the case that further tuning of the crossover and mutation rates is required, and alternative evolutionary operators such as expected crossover and mutation and ranked crossover are currently being explored. Also, the weighted sum method which aggregates all individual fitnesses within one overall, compensatory function is acknowledged as being sub-optimal in comparison to more sophisticated algorithms such as Shaffer's (1985, in Michalawicz 1992) VEGA system which rank populations according to pareto dominance (the ranked position according to each individual criterion, not accumulated fitness). the current formulation does not allow for the generation of two sets of sites for both testing and training. but this could be achieved in future work by the use of a double objective function.

CONCLUSIONS AND FURTHER DIRECTIONS

While simplistic in evolutionary computing terms, in comparison with traditional sampling methodologies the technique outlined appears rather complex. However, the approach outlined does allow for the sampling of fixed point sites of data both characteristic of the underlying landscape spaces and tailored for analysis. To date, only ad hoc alternatives exist. Additionally, with the tailoring of the criteria and therefore the fitness function, the methodology developed is equally applicable to the selection of other geographically referenced point data, such as for example soil survey results and indeed to very different geographical problems.

In order to facilitate the exploration of a wider variety of genetic operators and parameters, work is undergoing to visualise the problem search space according to the defined function as a means of guidance. The generation of equally representative training and testing site combinations in order to avoid over reliance on cross-validation techniques is also underway.

REFERENCES

Aspinall, R.J., Lees B.G. (1994) Sampling and analysis of spatial environmental data, In Waugh, T.C., Healey (Eds) (1994) *Advances in GIS Research, Proceedings of Sixth International Symposium on Spatial Data Handling*, Waugh: Edinburgh, p1086-1098.

Austin, M.P. , Heyligers, P.C.(1989) Vegetation survey design for conservation: Gradsect sampling of forests in north-eastern New South Wales, *Biological Conservation*, **50**, 13-32.

Bras, R.L., Rodriguez-Iturbe, I. (1976) Network design for the estimation of areal mean of rainfall events, *Water Resources Research*, **(12)6**, 1185-1195.

Carver, S. J. (1991) Integrating multicriteria evaluation with GIS, *International Journal of Geographical Information Systems*, **5**, 321-339.

Davis, L. (1991) *Handbook of Genetic Algorithms*, Van Nostrand Reinhold: New York.

Eastman, J., Jin, W., Kyem, P.A.K., Toledano, J. (1995) Raster procedures for multi-criteria / multi-objective decisions, *Photogrammetric Engineering & Remote Sensing*, **61(5)**, 539-547.

Fonseca, C.M., Fleming, P.J. (1995) An overview of evolutionary algorithms in multiobjective optimization, *Evolutionary Computing*, **3(1)**, 1-16.

Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley: Reading, Mass.

Jankowski, P. (1995) Integrating geographical information systems and multiple criteria decision-making methods, *International Journal of Geographical Information Systems*, **9(3)**, 251-273.

Janssen, R., Rietveld, P. (1990) Multicriteria analysis and GIS: an appliction to agricultural land use in the Netherlands, In *, Geographical Information Systems for Urban and Regional Planning*, Eds. Scholten, H.J., Stillwell, J.C.H., Kluwer : Dordrecht.

Lees B. (1993) Sampling strategies for machine learning using GIS, In *Proceedings of the Second International Conference on Environmental Modelling*, Breckenridge, NCGIA: CA.

Margules, C.R., Stein, J.L. (1989) Pattern in the distributions of species and the selection of nature reserves: an example from Eucalyptus forests in south-eastern New South Wales, *Biological Conservation*, **50(4)**, 219-38.

Michalewicz, M. (1992) *Genetic Algorithms + Data Structures = Evolution Programs*, Springer- Verlag: Berlin.

Pereira, A.G., Peckham, R.J., Antunus, M.P. (1993) GENET: A method to generate alternatives for facilities siting using genetic algorithms, In *European Conference on Geographical Information Systems*, 1993, 973-982.

Pettitt, A.N., McBratney, A.B. (1993) Sampling designs for estimating spatial variance components, *Applied Statistics*, **42(1)**, 185-209.

Ritzel, B.J., Eheart, J.W., Ranjithan, S. (1994) Using genetic algorithms to solve a multiple objective groundwater pollution containment problem, *Water Resources Research*, **30(5)**, 1589-1603.

Schwefel, H.-P., (1995) *Evolution and optimum seeking*, Wiley: New York.

Shaffer, J.D., Grefenstette, J.J. (1985) Multi-objective learning via genetic algorithms, In Joshi, A.(Ed), *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Morgan Kaufmann: San Mateo, CA, p593-595.

AFFILIATIONS

Ordnance Survey height data	Height
(1: 50,000 raster)	Second derivative of slope
	Aspect
* University of Edinburgh	Central Science Laboratory
Department of Geography	Ministry for Agriculture, Food & Fisheries

Drummond Street
EDINBURGH

Hatching Green
Harpenden
Herts.

M.F.Hutchinson

A LOCALLY ADAPTIVE APPROACH TO THE INTERPOLATION OF DIGITAL ELEVATION MODELS

Significant extensions to an existing finite difference approach to the interpolation of digital elevation models (DEMs) are discussed. The underlying finite difference computational structure is retained, but the minimum curvature interpolation criterion is replaced by a locally adaptive criterion which directly minimises profile curvature, that is curvature of the modelled land surface in the down slope direction. The new method should automatically match landforming processes and hence preserve drainage structure. It should remove biases in the representation of profile curvature which have previously limited applications to the analysis of erosion and soil forming processes. The local nature of the new interpolation condition should also permit greater flexibility in modelling landscape features, including breaklines and cliffs. A statistical analysis of the discretisation errors imposed by representing the landscape with a regular grid DEM leads to a method for optimal smoothing of both point and contour line elevation data. This in turn has led to a simple criterion for matching the spatial resolution of the DEM to the information content of the data, a practical advance toward addressing scale issues in hydrological and environmental modelling.

INTRODUCTION

Topography is a dominant control on earth surface processes. It moderates the spatial distribution of climate, an environmental variable of equal importance in controlling the distribution and productivity of biological systems. Topography also directly moderates the flow of water over and through the earth's surface, in turn moderating soil wetness and soil erosion potential, and influencing soil physical and chemical properties.

The chief limitations of regular grid elevation models (DEMs) appear to lie in not being adaptive to topography with spatially varying complexity and in supporting sometimes over simplistic hydrological analyses. The paper by Gallant and Hutchinson (1996) describes an alternative structural approach to digital elevation modelling. Despite these limitations, regular grid DEMs have become an increasingly popular way of encoding the topography for environmental modelling purposes. They are directly compatible with remotely sensed data sources and complex terrain can be represented provided the DEM has sufficiently fine resolution. This raises the issue of scale which has practical implications for data acquisition and storage and theoretical implications for dependent hydrological and other earth surface models. It also has implications for integration of DEMs with other spatial data sets.

The spatially varying complexity of actual topography and its strongly anisotropic (or directional) behaviour preclude a straightforward analytical approach to interpolation of elevation, especially in view of the very large data sets involved. Input and output data sets may contain over one million points. These difficulties can be addressed with analytical methods, by dividing the region into small adjoining pieces, but problems remain in deciding the size of such pieces, how to join them together and how to address anisotropic behaviour still within, and immediately beyond, each piece.

The finite difference approach developed by Hutchinson (1988,1989) has been shown to address many of these difficulties. Analytic functions are replaced by the regular grid of points and a high quality interpolation criterion related to minimum curvature is used to construct the basic interpolation method. The finite difference formulation can be made computationally efficient, by using a suitable multi-grid iteration method, and facilitates the incorporation of sensible drainage structure by imposing directional constraints on the grid points relating to the direction of surface water flow. The directional constraints can be inferred automatically by a drainage enforcement algorithm and obtained directly from digitised stream line data. The method can also infer directional constraints, defining ridges and stream lines, from the corners in fine scale contour data. With these extension, the method is thus already strongly locally adaptive, although the underlying minimum curvature algorithm is not. Two significant extensions to this finite difference methodology are discussed.

The first of these, which has been developed and already applied to the existing method, involves a statistical analysis of the discretisation errors imposed by representing the landscape with a regular grid DEM. This has led to a locally adaptive method for optimally smoothing both point and contour line elevation data. It has in turn led to the development of a simple criterion for matching the spatial resolution of the interpolated DEM to the information content of the source point and line data. This is intended to be a practical advance toward addressing scale issues in hydrological and environmental modelling.

The second extension, which is still under development, replaces the minimum curvature interpolation criterion by a locally adaptive criterion which directly minimises profile curvature, that is curvature of the modelled land surface in the down slope direction. The new method automatically matches landforming processes and hence preserves drainage structure. It holds the potential to remove biases in the representation of profile curvature which have previously limited applications to the analysis of erosion and soil forming processes. The local nature of the new interpolation condition should also permit greater flexibility in modelling landscape features, including breaklines and cliffs.

LOCALLY ADAPTIVE SMOOTHING OF DISCRETISATION ERROR

The current interpolation method employs a simple multi-grid approach which initially solves the iterative interpolation problem on a very coarse grid resolution. The problem is then solved at succeeding finer resolutions, until the final resolution is obtained. Starting values for each succeeding grid resolution are obtained by bilinear interpolation from the preceding coarser grid.

At each resolution the elevation data are simply allocated to the centre of nearest grid cell.

This introduces a small positional error in the data which may be interpreted as a vertical error in the point placed at the centre of the cell. This is illustrated in Figure 1 where the data point A on a sloping terrain surface is allocated to the centre of the grid cell of width h, leading to a horizontal displacement d, which gives rise to a vertical error z.

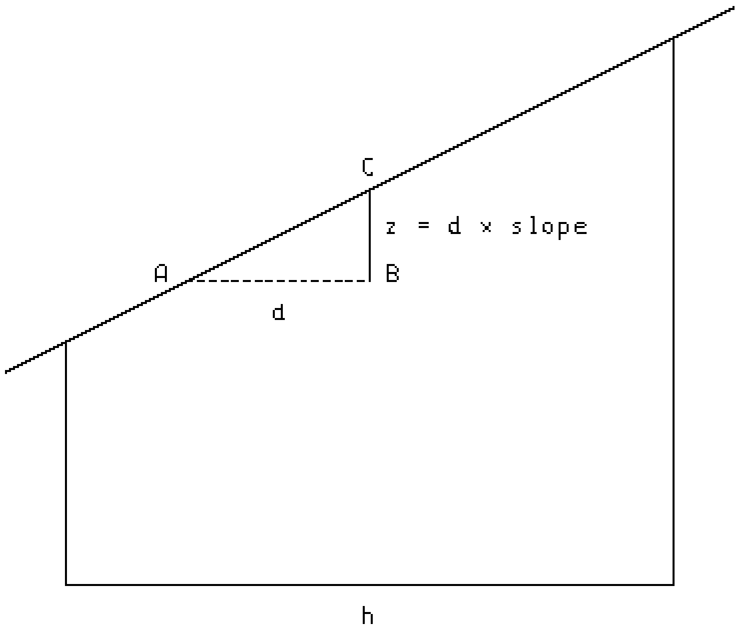


Figure 1. Vertical error of displaced data point.

The size of this vertical error depends on the slope of the grid cell and the size of the horizontal displacement as shown in Figure 1. If it is assumed that each data point is placed randomly within its associated grid cell then the variance of the corresponding vertical error can be shown to be

var = h*h*slope*slope/12.

The amount of data smoothing can then be determined by weighting the elevation data points, as allocated to the grid cell centres, according to their error variances as estimated by this equation, and determining the smoothing parameter so that the average of the squares of the weighted residuals is 1.0. This matches the expected value of the average of the weighted residuals of the displaced data points from the actual terrain surface.

The amount of smoothing is therefore locally adaptive to the slope of the terrain. Since slope is a function of the terrain surface, it cannot be specified until the terrain surface has been interpolated. The iterative nature of the interpolation procedure is crucial to the imposition of this locally adaptive weighting. It permits the use of steadily improving estimates of the terrain slopes until the process converges for each successive grid resolution. The elevation procedure ANUDEM has been upgraded to apply the locally adaptive data weighting. Care was required in adopting a suitably stable subsidiary iterative process which the smoothing parameter converges to a final value. Further details will be published elsewhere.

OPTIMISING DEM RESOLUTION

As the multi-grid iteration proceeds from coarse DEM resolution to the final specified resolution, the slopes of the fitted grid at the data points steadily increase in magnitude. At coarse resolutions several data points may be allocated single grid cells leading to averaging of the data and resultant smoothing of the fitted DEM in relation to the actual terrain surface. Eventually, the DEM resolution is sufficiently fine for there to be little or no data point averaging, and the slopes of the fitted DEM at the data points stabilise. At this stage, all information has been extracted from the source terrain data, since further refinement of the DEM resolution would not change the DEM slopes.

The African DEM (Hutchinson et al. 1996) was calculated at 10 grid spacings which were successively halved from 6.4 degrees to 0.0125 degrees. Figure 2 shows that the root mean square slope of the DEM at the data points stabilises to a value of 3.5% once the DEM resolution has reached the tenth grid resolution of 0.0125 degrees or about 1.25 km. The point data used in fitting this DEM were digitised from the 1:1M scale air navigation charts. The 1 km resolution used in applications by the USGS of the full 1:1M scale data would appear to be consistent with the information content of this source data set.

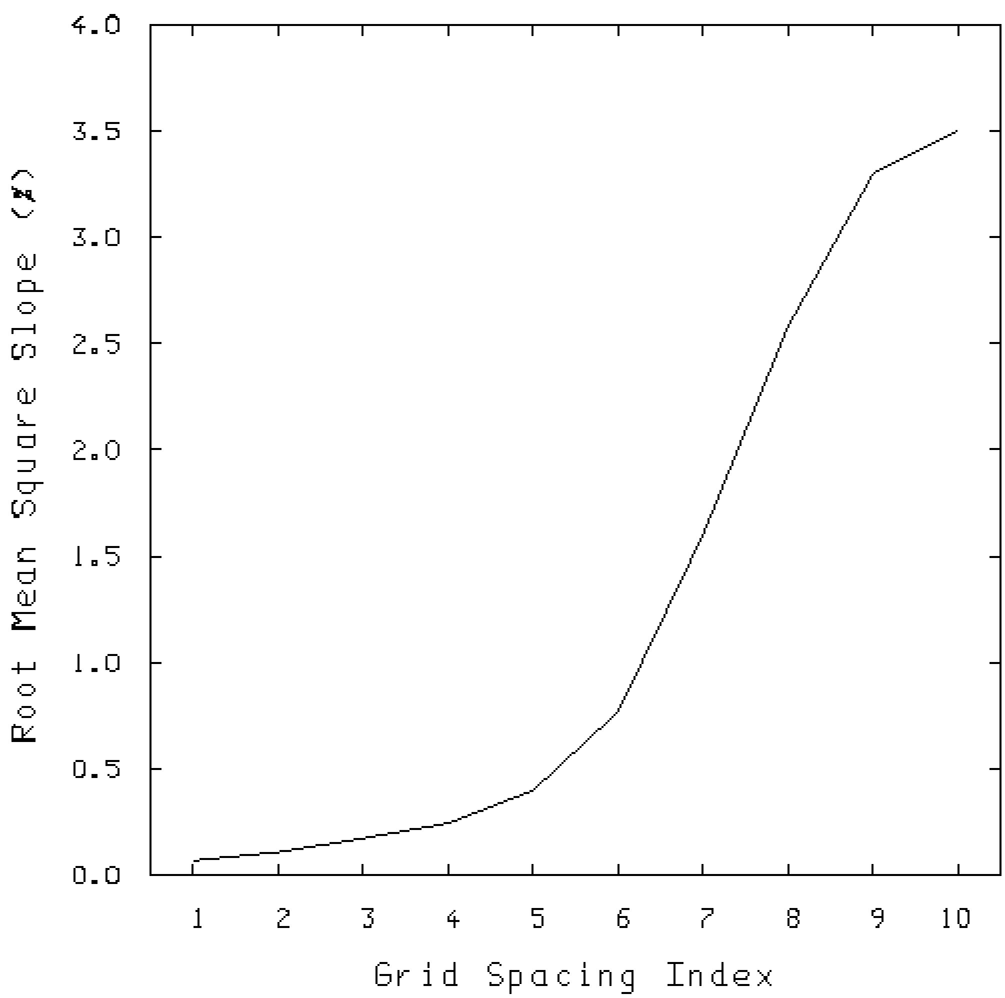


Figure 2. Root mean square slope vs resolution of African DEM.Figure 1.

This gives rise to a simple criterion for optimising DEM resolution to the information content of the source data. This has significant practical implications for data inventory, since data storage requirements are very sensitive to grid resolution. The smoothing of the discretisation error can be viewed as providing an optimal representation of the terrain surface at all resolutions up to the final optimised resolution. At each resolution the fitted DEM values are estimates of the average of the actual terrain surface across each grid cell. DEMs produced in this way have facilitated investigations by Gessler et al (1996) of the effects of DEM resolution on derived terrain parameters at various resolutions.

A LOCALLY ADAPTIVE INTERPOLATION CRITERION

The next step in the development of the locally adaptive method is to develop a locally adaptive roughness penalty to define the interpolation process. This has been achieved until now by applying various localised constraints to an underlying interpolation procedure defined by a general, non-adaptive, roughness penalty. It is intended that these local constraints be incorporated directly into the proposed adaptive roughness penalty. The proposed procedure is to minimise profile curvature of the DEM in the downslope direction.

The definition of this roughness penalty depends on the aspect of the terrain surface being interpolated. As for the adaptive data smoothing criterion, an iterative process is therefore required so that steadily improving estimates of terrain aspect are obtained as the iteration proceeds. It is anticipated that this process will be stable because terrain aspect is less sensitive to DEM resolution than profile slope and curvature. In fact, in the simple but reasonably typical situation of a hill with circular contours, it can be seen that the profile slope and curvature can be prescribed arbitrarily while aspect, as represented by the circular contours, remains constant.

A locally adaptive interpolation criterion also requires substantial additional computation. A prototype implementation has been developed which attempts to compensate for the additional computational burden by making the implementing most of the computations associated with each row of the DEM in a way which can be directly parallelised on a vector processor.

The stability of the overall process is a critical issue. In the current prototype additional stability has been imposed by adding a small factor times the original generalised curvature roughness penalty. An application to the small point elevation data set provided in Table 5.11 of Davis (1986) is shown in Figure 3. The roughness penalty used here was the sum of profile curvature and one tenth of the generalised curvature. The figure shows stronger identification of valley structure than the minimum curvature interpolation shown in figure 1 of Hutchinson (1989). The shape of the DEM is similar to that of the DEM shown in figure 6 of Hutchinson (1989), which was obtained by imposing specific drainage constraints on a modified minimum curvature interpolation procedure.

An preliminary application to contour data has reduced the bias towards the heights of the data contours exhibited by DEMs interpolated using non-adaptive roughness penalties. In particular the application has shown better extrapolation towards peaks defined by surrounding contour line data. This should also reduce biases in representation of profile curvature, as the bias towards the data contour heights using the general roughnes penalty is due to a slight flattening of the fitted DEM in the vicinity of the data contours.

There appears to be trade-off between stability, as currently obtained by adding a generalised curvature roughness penalty, and sensitivity to drainage structure. If the generalised curvature factor is too small then aspect is not stably defined. On the other hand, if it is too large, then drainage structure is too generalised. Further developments in the algorithm are anticipated.

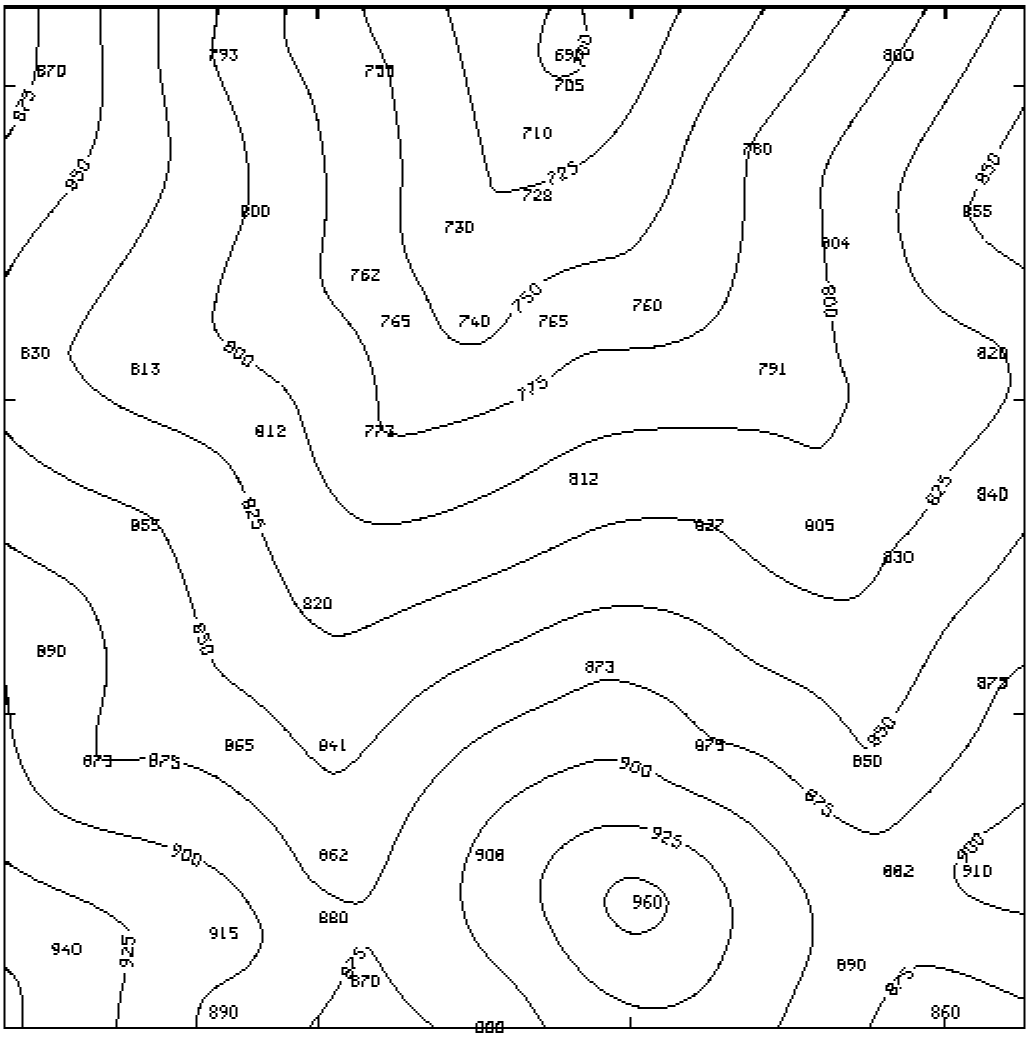


Figure 3. Locally adaptive interpolation of Davis elevation data.

CONCLUSION

The locally adaptive approach to digital elevation modelling shows promise in matching the spatially variable complexity of actual terrain surfaces. Locally adaptive smoothing of discretisation error has been successfully implemented and has yielded a useful criterion for optimising resolution of fitted DEM to the information content of the source data.

The locally adaptive roughness penalty also shows promise, but problems associated with maintaining stable and accurate definition of terrain aspect during the interpolation process need to be satisfactorily resolved. The parallelisation of the computations helps to reduce the additional computational cost of the locally adaptive technique.

REFERENCES

Davis,J.C. 1986. *Statistics and Data Analysis in Geology*, *Second Edition,* Wiley, New York.

Gallant,J.C. and Hutchinson,M.F. 1996. Towards an understanding of landscape scale and structure. These proceedings.

Gessler,P.E., McKenzie,N. and Hutchinson,M.F. 1996. Progress in soil-landscape modelling and spatial prediction of soil attributes for environmental models. These proceedings.

Hutchinson,M.F. 1988. Calculation of hydrologically sound digital elevation models. *Proceedings, Third International Symposium on Spatial Data Handling, Sydney,* Columbus: International Geographical Union, pp. 117-133.

Hutchinson,M.F. 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology* 106: 211-232.

AUTHOR INFORMATION

M.F.Hutchinson
Centre for Resource and Environmental Studies
Australian National University
Canberra ACT 0200
Australia
email: hutch@cres.anu.edu.au

Kristine L. Verdin and Susan K. Jenson

Development of Continental Scale Digital Elevation Models and Extraction of Hydrographic Features

---

Introduction

Continental scale digital elevation models (DEMs) are required for climate and global change studies spanning a variety of disciplines: atmospheric science, hydrology, biogeochemistry, wildlife biology, forestry, range science, and others. To date the only available data set with global coverage has been ETOPO5 (National Geophysical Data Center, 1988) which, with its 5 arc-minute (approximately 10 km) resolution, has proven inadequate for many applications.

To better address the requirements of the scientific community, the U.S. Geological Survey (USGS) has undertaken to provide DEM data sets for all the Earth's land masses at 30 arc-second (approximately 1 km) resolution. USGS scientists are working with the Defense Mapping Agency (DMA) to develop procedures for generalizing 3 arc-second (approximately 90 meter) digital terrain elevation data (DTED 1) for large land masses. For areas where DTED 1 is not available, DMA's Digital Chart of the World (DCW) is being used as a source to generate elevation data, using a gridding algorithm that captures contours, spot heights, and hydrology with drainage enforcement. Techniques were developed to blend the DCW-derived DEM data with the generalized DTED. To date (January 1996), Africa, North America, Europe and Asia are complete and are being made available, free of charge, to the public via internet transfer. A 30 arc-second DEM of Antarctica is being finalized with distribution anticipated for early 1996. Work is progressing on the continents of South America and Australia, with an anticipated completion date of October 1996.

In addition to providing the raw elevation data, the DEMs are being processed to extract hydrographic features. The North American DEM was used as a prototype to demonstrate the feasibility of extracting basins and flowlines from the USGS 30 arc-second DEMs. Derived basins and flowlines were verified with independently mapped source material.

Description of Products

The USGS global 30 arc-second DEM project is being carried out at the USGS EROS Data Center in Sioux Falls, S. Dak. The broad goal of this project is the completion of global 1 km elevation data for the land surface and the systematic extraction of derivative information to assemble a global data base of topographic elevation, slope, aspect, hydrologic flow paths, and watersheds. The primary collaborators and data contributors to date are the USGS, DMA, NOAA, NASA, UNEP/GRID, USAID, the Instituto Nacional De Estadistica Geografia E

Informatica (INEGI) of Mexico, and the Geographical Survey Institute of Japan.

The first data set completed for the project was of the African continent, developed before permission was obtained to incorporate the generalized DTED1 data. This data set is entirely derived from gridded DCW vector data, but work is progressing on incorporating the generalized DTED1 data. The first data set that was developed using the two different data sources was of the North American continent. This data set incorporates generalized 3 arc-second data from various sources, along with gridded DCW data. Techniques were developed with the North American data set to blend the various sources of data to minimize discontinuities between the data types. The European and Asian data sets were developed using the same techniques, as will be the South American DEM. Antarctica is being developed entirely from gridded vector product because DTED1 data do not exist for Antarctica.

The DEMs are being distributed as compressed images, 16-bit binary raster image files in a latitude/longitude coordinate system. The elevation values are given in meters above mean sea level with ocean areas masked as -9999 (representing no data). Inland water bodies carry representative surface elevations. Vertical accuracies of the elevation values vary with the data source. They range from root mean square error (RMSE) of 18 m to about 90 m (Gesch, 1994). The DEMs are currently available by ftp transfer from the EROS Data Center anonymous ftp site or tape request. Plans for CD-ROM distribution of the DEMs and ancillary data sets are proceeding. The data sets consist of files of z values, corresponding to geographic subsets of the continental coverage areas, plus 4 ancillary files created with ARC/INFO GIS software (header, world, tic, and statistics files) and the Land Analysis System image processing software (DDR file)along with README documentation files.

Input Data Sources

The 30 arc-second elevation values were assembled from various data sources. While the African data set is currently a product of the gridded DCW vector product only, the other continental data sets currently available are derived from different sources. Figures 1, 2 and 3 depict the sources used for the North American (figure 1), European (figure 2) and Asian (figure 3) data sets. The sources can be characterized as essentially two types:

1. Existing digital elevation data at 3 arc-second spacings (approximately 90 m) were obtained wherever possible and generalized to the desired 30 arc-second cell size. Various sources were used to obtain the data. These data provided coverage for the conterminous United States, Alaska, Mexico, Europe, Asia and parts of Canada, Central America, and the Caribbean Islands. These data account for 60 percent of the data incorporated into the North American DEM, 99 percent of the European DEM, 50 percent of the South American DEM and 5 percent of the Asian 30 arc-second DEM.
2. The DCW 1:1,000,000 mapping (Defense Mapping Agency, 1989) and various other vector data sets were used to interpolate elevation values in the areas not covered by DTED1. The ANUDEM splining algorithm (Hutchinson, 1989), which utilizes mapped hydrography along with vector and point hypsography to produce a hydrologically correct DEM, was used.

The continental data sets are being distributed as they are completed. Coverage of the entire world at 30 arc-second resolution is targeted for completion by the end of 1996. The DEM data sets that have been completed and are currently available to the public are North America (figure 4) , Africa (figure 5) , Europe (figure 6) and Asia (figure 7) . The data currently are staged and available for free downloading from the EROS Data Center anonymous ftp site.

Drainage Analysis

Drainage analysis software (Jenson and Domingue, 1988; ESRI, 1992) is being employed to derive river basin data sets from the 30 arc-second DEMs. These tools are based upon an algorithm that solves for cell to cell flow directions by finding the path of steepest descent between a cell and its eight neighbors. Thereafter, values for flow accumulation (the number of tributary cells) are computed for each cell, permitting the definition of drainage networks of varying density, and, for any cell of interest, the identification of the area tributary to it.

An abundance of data covering the North American continent can be used to verify the basins derived from the 30 arc-second DEM. Foremost among these data sources is the Hydrologic Unit System developed by the USGS (Seaber, 1987). This system divides the entire United States into 21 major regions, 18 within the conterminous United States. These regions are further subdivided into 222 subregions including areas drained by a river system, a reach of a river and its tributaries in that reach, closed basins or a group of streams forming a coastal drainage area. Beyond these subregions the areas are broken into successively smaller accounting units and cataloging units. The shapes of accounting and cataloging units are often influenced by planning or administrative boundaries, whereas the region and subregion divisions are, in general, topographically delineated. An 8-digit code, commonly referred to as the Hydrologic Unit Code (HUC), is used to allow each of the four levels of classification to be uniquely identified by 2-digit fields within the 8-digit number.

To create a river basin data set for North America, drainage analysis software was applied to produce basin delineations for comparison with existing USGS vector coverages for the 18 regions (2-digit HUC) and 222 subregions (4- digit HUC) of the conterminous United States. These were compiled and digitized at 1:250,000 scale (Seaber, 1987), and are distributed to the public as a digital line graph (DLG) data set. Upstream areas were derived from the 30 arc-second DEM for the cells corresponding to the mouths of these basins. Results were then overlaid with the vector HUCs for visual inspection and tributary areas calculated for comparison.

Basin Delineation

Development of major basins from a continental-scale DEM presented challenges in the application of the existing drainage analysis algorithms. A DEM of the size of the North American 30 arc-second DEM contains many sinks, locations in the DEM from which there does not exist a natural drainage path. The vast majority of these sinks are spurious, resulting from the methodology used to develop the original DEM or introduced in the generalization algorithm. Before proper flowlines and resulting basins can be extracted from the DEM, these spurious sinks must be removed. The drainage analysis tools developed by Jenson and

Domingue (1988) and implemented in ARC/INFO provide the means to remove these spurious sinks by using a filling function. However, due to the complexity of the landscape on a continental scale, not all sinks occuring in the DEM are, in fact, spurious. Many sink features in the DEM are representations of natural surface features. Exclusion of these features by blindly filling all sinks will result in a DEM that will not correctly represent actual topography. The Great Basin of the western United States is a case in point. This large basin, one of the 18 major regions of the conterminous United States, is entirely closed; that is, there is no natural outlet to the sea. Blindly applying filling algorithms to the North American DEM would result in the Great Basin being filled and drained to the Pacific Northwest.

The North American 30 arc-second DEM was processed using standard GIS software tools to create a DEM from which major drainage basins can be delineated. Application of the existing tools in the GIS, such as filling, flow accumulation, and watershed functions, was done on the entire data set with techniques developed to ensure the preservation of actual sink features. The procedures developed for delineating basins for the North American continent can be summarized as:

- Project DEM to an equal area projection (Lambert Azimuthal Equal Area was used for North America).
- Use standard filling functions to fill the entire DEM without consideration of maintaining any sinks.
- Determine areas of "sinks" in DEM by differencing the original and filled DEMs.
- Create a mask of sink areas, group into uniquely numbered zones and calculate areal extent of sinks and maximum depth.
- Threshold sink zones according to area or depth criteria. The sinks in the North American DEM were thresholded by area, with sinks with areal extent in excess of 10,000 sq. km. being considered for maintenance.
- Check sink areas and discard those that correspond to reservoirs and other non-sink features.
- Place a NODATA seed at minimum elevation in each sink zone.
- Fill the DEM, check, and iterate.

Several iterations were carried out with the North American DEM before the thresholding criterion for the sinks was properly identified. The unique situation of the Great Lakes region in the data set also required some iterating to ensure that the flow through the Great Lake system was properly modeled.

Results

Figure 8 presents a comparison of the 2-digit HUCs derived from the 30 arc-second DEM, shown as colored polygons, and the boundaries of the 2-digit HUCs as recorded in the 1:250,000 DLG product, shown as black lines.

Disagreements are seen to be minor relative to the extent of the continental area that has been subdivided. Figure 9 highlights the discrepancy areas as small colored polygons along the divides between the 2-digit HUCs.

Most of these coincide with areas of low relief, and are explained by the inadequacy of a 30 arc-second elevation spacing for resolving the subtle features that control drainage in those areas. An exception to this observation is the Riviere Richelieu, which the DLG product includes as part of the Mid-Atlantic Basin for administrative purposes. This basin, in fact, drains into the St. Lawrence River, and is therefore part of the Great Lakes Basin, which drainage analysis of the 30 arc-second DEM has correctly delineated.

Table 1 presents a comparison of the two data sets by listing the respective areas for the 2-digit HUCs as represented by the DLG and DEM-derived data sets. Differences are seen to be typically less than 5 percent of the areas in question, except for the Lower Mississippi and Mid-Atlantic Basins, already mentioned, and New England.

Huc #	Region Name	HUC area (sq.mi.)	HUC area (sq.km.)	DEM generated (sq.km.)	Discrepancy (%)
1	New England	64090	165990	155280	-6.5%
2	Mid-Atlantic	111360	288420	255470	-11.4%
3	South Atlantic-Gulf	278680	721780	697300	-3.4%
4	Great Lakes	178300	461800	482730	+4.5%
5	Ohio	161250	417640	411650	-1.4%
6	Tennessee	40670	105340	105800	+0.4%
7	Upper Mississippi	189100	489770	491940	+0.4%
8	Lower Mississippi	104030	269440	255060	-5.3%
9	Souris-Red-Rainy	60350	156310	162370	+3.9%
10	Missouri	509547	1319730	1308270	-0.9%
11	Arkansas-White-Red	245500	635840	650460	+2.3%
12	Texas-Gulf	183140	474330	461160	-2.8%
13	Rio Grande	132510	343200	344810	+0.5%
14	Upper Colorado	112110	290360	293790	+1.2%
15	Lower Colorado	139130	360350	372120	+3.3%
16	Great Basin	140110	362880	372540	+2.7%
17	Pacific Northwest	277660	719140	712930	-0.9%
18	California	159650	413490	397400	-3.9%

Table 1. Comparison of Basin Areas

Development of prototype basin extraction procedures for 30 arc-second DEMs for the North American case permitted checking results against existing digital depictions of basin boundaries for the United States, as described, and for Canada. (Most other continents are without these valuable reference data sets.) The experience revealed that it is necessary to work closely with sink areas in the DEMs, to discriminate between spurious sinks and natural closed basins. Even so, the results favor the application of the approach to all continents to obtain a consistent and comprehensive global basins data set. Discrepancies with delineations that can be obtained with higher resolution elevation data are not significant at global scales.

When completed, the USGS global basins data set will be a valuable resource for partitioning results obtained by observing and modeling atmospheric and hydrologic processes. This will facilitate the calculation of meaningful continental and regional energy and water balances.

SELECTED REFERENCES

Defense Mapping Agency, 1989, Digital Chart of the World Database (MIL-D-89009), Washington, D.C., U.S. Government Printing Office.

Defense Mapping Agency, 1990, Digitizing the future (3d ed.): Defense Mapping Agency, Washington, D.C., 105 p.

Defense Mapping Agency, 1992, Development of the Digital Chart of the World: Washington, D.C., U.S. Government Printing Office.

ESRI, 1992, "Cell Based Modeling with GRID", ESRI, Inc.

Gesch, Dean B., 1994, "Topographic Data Requirements for EOS Global Change Research", U.S. Geological Survey, Open-File Report 94-626, 60 p.

Hutchinson, M.F., 1989, A new method for gridding elevation and stream line data with automatic removal of pits: J. Hydrol, 106, 211-232 p.

Jenson, S.K., "Applications of Hydrologic Information Automatically Extracted From Digital Elevation Models," Hydrological Processes, 5:1, 31-44 (1991)

Jenson, S.K. and Domingue, J.O., 1988, "Extracting topographic structure from digital elevation data for geographic information system analysis: Photogrammetric Engineering and Remote Sensing, v. 54, p. 1,593-1,600.

National Geophysical Data Center, 1988, Topography data base - data announcement 88-SE-1102: National Geophysical Data Center, Boulder, Colo.

Seaber, Paul R., F. Paul Kapinos, and George L. Knapp, 1987, Hydrologic Unit Maps, United States Geological Survey Water-Supply Paper 2294, 63 p.

AUTHOR INFORMATION

Kristine L. Verdin
Hughes STX
EROS Data Center
Sioux Falls, SD 57198
(605) 594-6002

email: kverdin@edcsgw6.cr.usgs.gov

David R Miller and Jane G Morrice

Assessing uncertainty in catchment boundary delimitation

This paper reports on the issues of delimiting water catchments using topographic maps in digital or graphical form and their implications in catchment based modelling. An assessment is presented of the derivation of catchment boundaries from sources of different scales and the representation of boundary reliability. A simulation of error in the DEM is used to recalculate the watershed boundaries and assess the stability of their location with respect to the elevation data used. Finally, the assessment of catchment boundaries is applied to a database of catchments at a national level.

INTRODUCTION

This paper presents an assessment of the uncertainty associated with delimiting catchment boundaries of water catchments and the representation of this within a Geographic Information System (GIS). The scale of input data, information content and the representation of boundary reliability were studied and a model of the reliability of catchment delimitation are presented.

The study of environmental applications with respect to water catchments is due to their functional role in integrating the chemical, biological and physical processes within their boundary. Any process which relates to water runoff or terrain slopes are, by definition, operating within a water catchment of some form (Moore, *et al*, 1993). The use of DEMs as a basis for deriving secondary products such as slope or aspect, exposure or water catchments and intervisibility, has been well documented by several authors (Maidment, 1993; Moor, 1993; Fisher, 1994). Data and data quality were discussed in terms of the inherent errors in topographic data and the representation of topography in digital form (Clark, 1993; Fisher, 1994; Goodchild & Han, 1995). The conclusions have been that DEMs and the products derived from them deserve critical appraisal but that their use was still dominated by what has been available.

Despite the considerable amount of work undertaken on hydrological applications within a GIS there has been relatively little on the limitations of automating the process of catchment delimitation, how such delimitation compares to a manual delimitation (Avissar, 1993; Merkel & Sperling, 1993) and the sensitivity of automated techniques to errors in digital elevation models (DEMs).

This paper reports on a comparison of the graphical delimitation of catchment boundaries at a range of map scales available for Great Britain, chosen because of their current use in environmental applications such as water runoff modelling and sewage sludge disposal (Beven & Wood, 1994; Klaghofer *et al*, 1993; Towers, 1994). A digital delimitation has been undertaken for 1:50 000 scale digital elevation data and re-derived based on perturbations of the elevation data and an assessment of the stability of the boundaries.

MANUAL DELIMITATION

A study area was selected to compare manual delimitations of catchment boundaries. The area selected was the Trossachs in central Scotland which is approximately 50 km north of Glasgow on the eastern side of Loch Lomond (Figure 1). The topography ranges from approximately 50 m to 880 m above Mean Sea Level, with slopes of between level and 55o plus cliff faces. The catchment boundaries (or watersheds) within this area vary from those with shallow slopes to watersheds with steep slopes on one or both sides running towards different lochs.

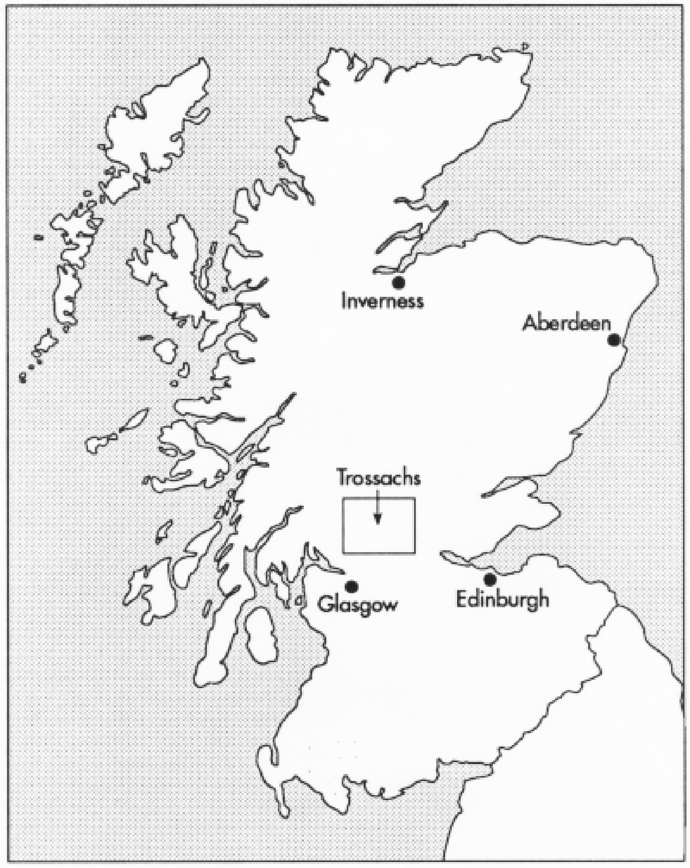


Figure 1. Location of study area

Three scales of topographic maps were used in this study: 1:50 000, 1:25 000 and 1:10 000. Each of the catchments for the eleven lochs were delimited using the topographic maps as a base and digitizing for storage and analysis in the ARC/INFO GIS. The catchment boundaries were delimited, by a combination of interpretation of height as presented on the map using contour lines, trigonometric pillars and spot heights plus features depicting surface hydrology, such as rivers and lochs. Figure 2 illustrates the catchment boundaries.

The cumulative effect of the digitizing reference points and adjusting between the local reference system (the digitizing tablet) and the National Grid on the locational accuracy of a point being digitized on a line 0.5 mm thick is +-0.6 mm at map scale (Mikhail & Gracie, 1981; Chrisman, 1983). This equates to approximately +-6 m at ground scale using a 1:10 000 scale map, +-15 m at 1:25 000 and +-30 m at 1:50 000 scale and these values have been used to represent the error bands around each boundary segment due to the cartographic process (Chrisman, 1983).

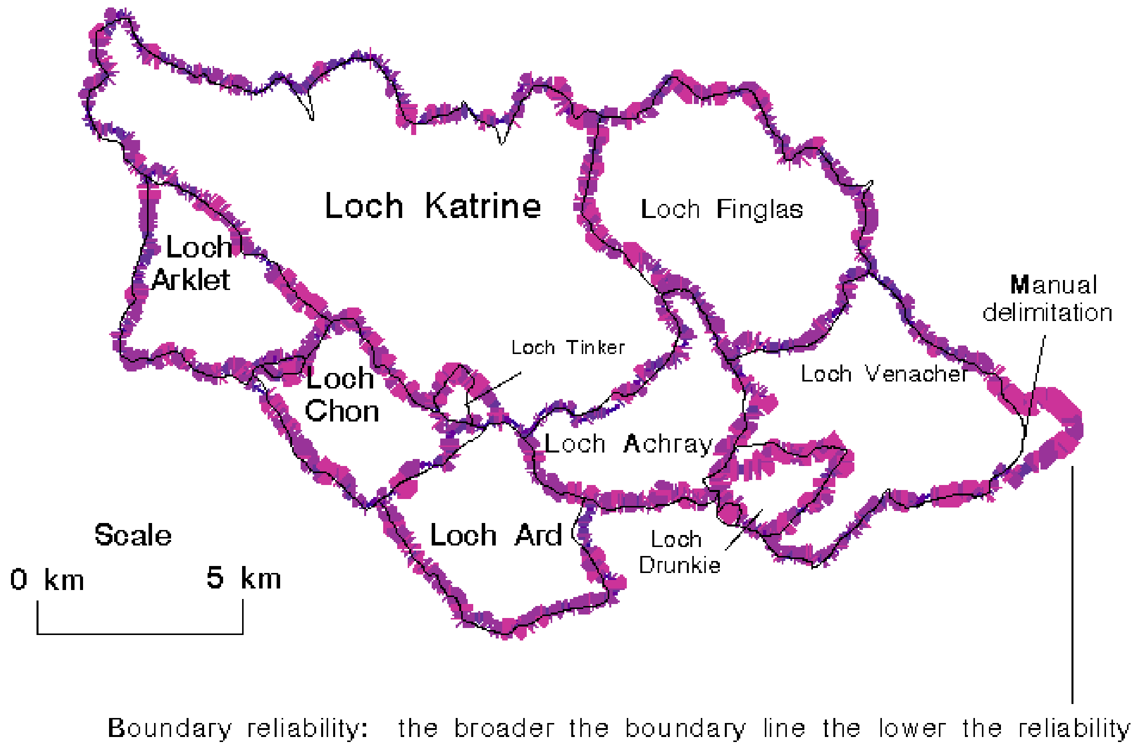


Figure 2. Representation of catchment boundary uncertainty

The total perimeter of catchment boundaries in the area is 168 km of which the spread in the locations of the three boundaries is greater than 200 m for less than 5% of the total (a maximum of 780 m). The spread over 70% of the perimeter is less than 25 m, which less than the level of accuracy expected by the cartographic procedures used. However, for the four smallest catchments the percentage of the perimeter that varies by over 200 m is up to 35% (for Loch Drunkie). Overall, the boundaries derived from the 1:10 000 and 1:25 000 scale map data are in greatest agreement compared to the 1:50 000 data. The difference between boundaries is less than 25 m for over 85% of the length of the boundary.

The impact of a re-definition of the boundary between catchments resulted in a distribution in the spread of area estimates around the catchments in a pattern unrelated to that of catchment size. Thus, the boundary which has the highest percentage of its length within the bounds of cartographic uncertainty between the two map scales (78%) is that between Loch Achray and Loch Ard which contributes to a total of 0.08 km2 of land which may be allocated to one catchment or its neighbour. Whereas, the location of the boundary between Lochs Drunkie and Venachar, which has the lowest level of agreement (32%), contributes 0.9 km2 to the range in areas between Lochs Drunkie and Achray and Drunkie and Lochan Reoidhte.

The interpretation of catchment boundaries using topographic maps depends upon the depiction of both altitude and water features. The level of detail of both aspects of the topographic map are dependent upon the scale of mapping (both source and published scales) and the compilation guidelines of the mapping organization. The compilation of the 1:25 000 map is based upon the 1:10 000 map therefore, the difference in details represented is due to cartographic generalization. The 1:50 000 maps are derived as separate products and thus the photogrammetric mapping and photographic interpretation are independent of that for the larger scale maps. Further, the contour interval for the larger scale maps is 5 m, whereas it is 10 m at the smaller scale. A consequence of this for interpreting catchment boundaries manually is that the level of information available to the interpreter is not the same for each map. The consequences are exemplified in the difference between the boundary locations for Loch Drunkie, where the additional contours produced different boundary interpretations.

AUTOMATED DELIMITATION

The catchment boundaries were delimited using an algorithm based upon the calculation of contributing upslope areas (ESRI, 1994; Jensen, 1991) and a 1:50 000 scale DEM (Ordnance Survey, 1995).

To ascertain the sensitivity of catchment boundary interpretation to topographic variation, the rates of change of altitude and aspect were calculated against which the interpreted boundaries were compared. The rate of change of height and aspect are effectively both representations of ridge lines, along which many catchment (or sub-catchment) boundaries will follow. From these datasets, each cell (in a raster representation) of the catchment boundary can be coded to represent the local terrain conditions and thus the relative levels of reliability of the catchment boundary at that location (Figure 1). This indicate where, along the boundary, it may be particularly unreliable rather than providing absolute estimates of the where the boundary may lie.

Comparing boundary locations, the average perpendicular distance between the two 1:50 000 source scale boundaries was approximately 65 m, ranging from 0 m to 1425 m (at the east end of Loch Venachar). The largest difference in locational position between catchment boundary intersections is approximately 940 m, at the boundary between Lochs Tinker, Chon and Ard. In this situation the manual delimitation has taken evidence of surface streams into account whereas the automated technique does not and thus has lead to delimitation of a boundary in a different location, re-allocating 0.6 km2 of land between catchments.

IMPACT OF DEM ERRORS ON DELIMITED CATCHMENT BOUNDARIES

The derivation of the DEM, the base data and the presentation of DEM data in a raster form provide sources of uncertainty in the product (Grayson *et al*, 1993; Rieger, 1993). The implication of locational errors is to mis--represent the absolute height at any particular location. Therefore, an error in location in the Eastings and Northings plane will have a consequent impact on the height value at each point. Such errors are highly spatially correlated (Fisher, 1994; Hutchison and Gessler, 1994)

Alternative representations of the elevation model have been used for assessing the sensitivity of viewsheds to errors in elevation data (Fisher, 1994; Monkton, 1994). The approach has used a standard deviation of 3.0 m for the Ordnance Survey 1:50 000 DEM, within which the elevation cells have been perturbed according to a normal distribution. The same approach has been employed in this study, for calculating alternative representations of the surface. After deriving a "new" realization of the elevation model, the DEM was scanned to identify all "sinks" and "peaks" before recalculating the catchment boundaries. This prevented inconsistency in the elevation model with respect to the direction of flow which would result in (small) spurious catchments being defined. The sensitivity of the derived boundaries to the removal of such sinks and peaks has not been assessed.

Figure 3 illustrates those areas of land that would be re-allocated between catchments with changes in the elevation model of different magnitudes. A total of 5.2 km2 would be re-allocated of which 59% would be caused with a change of less than 1 sd, 26% with changes of less than 2 sd and 11% with changes of 2.5 sd in the height model. Unfortunately, the accuracy of the DEM is not known locally and therefore, the estimation used is a global one. However, it is indicative of where catchment boundaries would be susceptible to uncertainty in the elevation data.

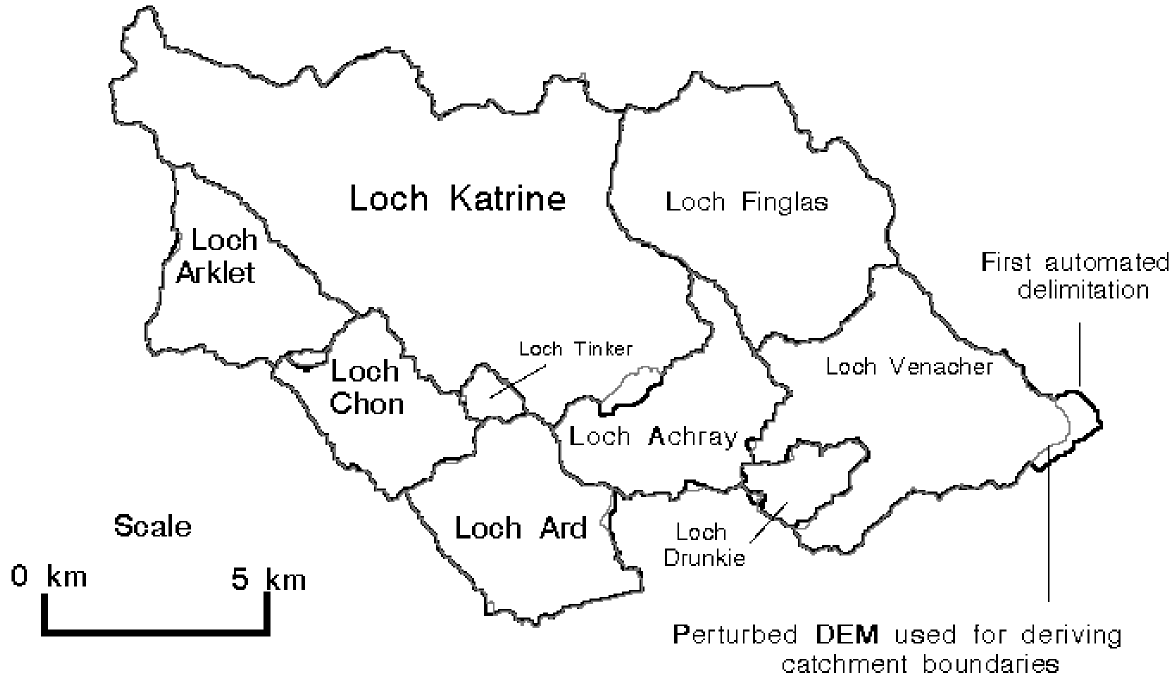


Figure 3. Delimitation of catchment boundaries based upon a perturbation of the digital elevation model

The results of this analysis indicates that some of the area (82%) of land identified by human interpretation would be correctly allocated with changes in altitude of less than 1 sd. This will not be the case where surface drainage runs across slopes in incised courses or is ditching. However, if the difference is due to minor perturbations in the DEM surface in areas of relatively even surface, the added area can be used to target land where closer examination of the catchment boundary may be warranted.

IMPLEMENTATION OF DELIMITATION ASSESSMENT ON A NATIONAL DATASET

A national database of 700 standing waters in Scotland has been compiled with their associated catchments delimited manually and various environmental datasets have been used to characterize the catchments, including land cover, soils, rainfall, evapo-transpiration and the river networks. Whole catchment modelling of the risks of acidification and eutrophication have employed such environmental characteristics and associated attributes to classes, such as phosphate loss to land cover classes (ref).

Whole catchment modelling generally requires assessments of such characteristics to be expressed in terms of proportions of the catchment from which contributing inputs of, for example, pollutants may be derived for each catchment. In Scotland the contibuting waters to lochs are predominantly surface rather than basin supplies. Uncertainties associated with the modelled outputs will be related to, in part, the uncertainties associated with the location of the delimited boundary which will impact on the estimated areas of land cover or soil types or the collecting area of rainfall.

The model of catchment boundary stability has been applied to the national dataset and the associated boundaries. One measure of the sensitivity of the catchments to the uncertainty in boundary location is to express the area of land that is subject to re-allocation due to uncertainty in the DEM as a proportion of the catchment area, either derived manually or automatically. The former representation is one of stability of method compared to input data and the second is of impact of data compared to a result with the added information obtainable from the map, and the users' most likely current starting point. Figure 4 illustrates the sensitivity of the catchment in the national database with respect to errors in the digital elevation model and thus the stability of the catchment boundaries.

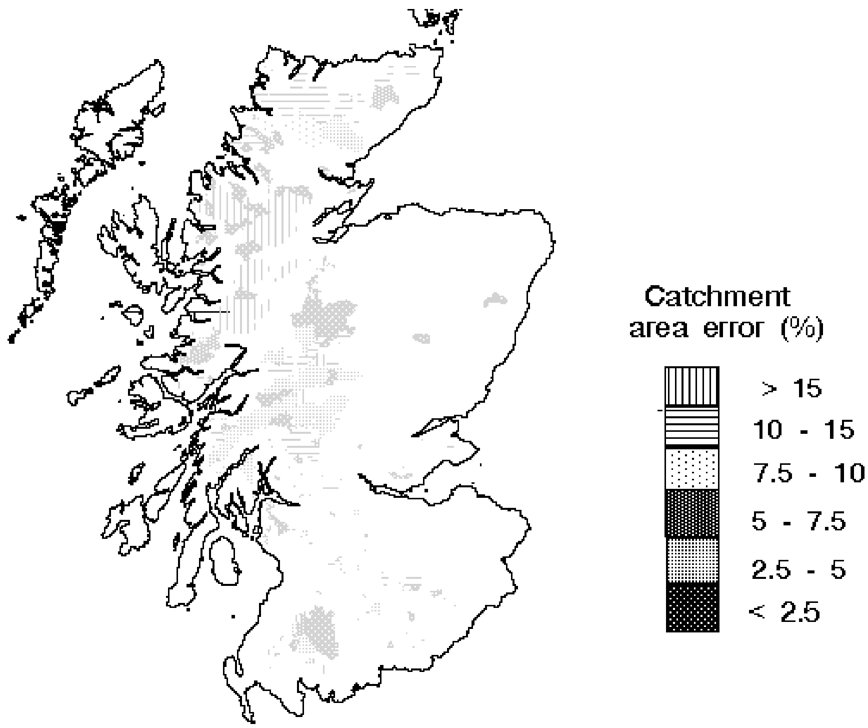


Figure 4. Water catchment sensitivity to errors in digital elevation data for boundary delimitation.

DISCUSSION

The extent to which the magnitude of the difference between automated delimitation an manual delimitation are significant is largely dependent upon the use to which the catchment boundary data is intended. Any application at a smaller scale than that at which the delimitation took place may not be sensitive to the differences in catchment boundary location and area described in this paper.

The use of automated techniques will increase as the availability of data increases and the implications of using one scale of source data compared to another will become more apparent. Manual techniques will remain important, at least to check the logic of the results of automated approaches. However, with more complete versions of vector data becoming available as standard products by national agencies the need for a more rigorous approach to automated delimitation may become greater because more users will encounter inconsistencies between the derived boundary and overlaid hydrological networks. The quality of the elevation data itself will also come under more rigorous appraisal. The consequences of not identifying erroneous peaks in the DEM are that streams which would have been automatically derived using hydrological modelling software will remain undefined. In the meantime, the DEM in the proximity of each sink has to be treated on a case by case basis. Under certain circumstances (such as the false interpolation) the sinks should be filled. However, on occasions when the problem is caused by presence of a peak, the peak needs removed rather than the sink infilled.

An analysis of individual boundary segments with respect to each map scale provides a basis for assessing the sensitivity of catchment areas according to each component segment that comprises the catchment's boundary. Therefore, one could identify the areas for which a larger scale of source data may improve the reliability of the definition of the catchment boundary. The same principal may be applied to boundaries derived from DEMs, except in the latter case one could identify groups of cells (in a gridded data model) for which a larger scale of source data may be of value. Thus, the boundaries could be constructed, piecemeal, from different scales of source map. In conclusion, the use of a conventional algorithm with digital elevation data and the veracity of the results will increase attention on the data used and secondly on the nature of the algorithms employed.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Ordnance Survey for the use of the digital elevation model and funding from the Scottish Office Agriculture and Fisheries Department.

REFERENCES

Aspinall, R J & Pearson, D (1995) Describing and managing data quality for categorical maps in GIS. In: *Innovations in GIS 2,* (ed. by P. F. Fisher), Taylor and Francis.

Avissar, R (1993) Relevance of Geographic Information Systems of landscape characteristics for hydroclimatological models. In: *Application of Geographic Information Systems in Hydrology and Water Resources Management* (ed. by K. Kovar & H. P. Nachtnebel), IAHS Publication 211, 75-82.

Band, L.E. (1989) A terrain-based watershed information system. *Hydrological Processes,* 3, 151-162.

Beven, K J & Wood, E F (1994) Catchment geomorphology and the dynamics of contributing areas. *Journal of Hydrology*. 65, 139-158.

Chrisman, N R (1983) Epsilon filtering: a technique for automated scale changing. In: *Proceedings of ACSM Annual Meeting,* 322-331.

Clark, M J (1993) Data constraints on GIS application development. In: *Application of Geographic Information Systems in Hydrology and Water Resources Management* (ed. by K. Kovar & H. P. Nachtnebel), IAHS Publication 211, 451-463.

ESRI Ltd., (1994) *GRID Reference Manual. ESRI Ltd.,* Palm Springs, USA. pp. 57.

Fisher, P.F (1994) Probable and fuzzy models of the viewshed operation. In:*Innovations in GIS 1,* (ed. by M. F. Worboys) Taylor and Francis, 167-176.

Goodchild, M F & Han, X (1995) The effects of topographic error in GIS. *International Journal of Geographic Information Systems.* 9 (2).

Grayson, R B, Bloshcl, G, Barling, R D and Moore, I D (1993), Process, scale and constraints to hydrological modelling, In: *Application of Geographic Information Systems in Hydrology and Water Resources Management* (ed. by K Kovar and H P Nachtnebel), IAHS Publication 211, pp. 83 - 92.

Jensen, S K (1991) Application of hydrologic information automatically extracted from digital elevation models. *Hydrological Processes*, 5 (1), 31-44.

Klaghofer, E, Birnbaum & Summer, W (1993) Linking sediment and nutrient export models with a geographic information system. In:*Application of Geographic Information Systems in Hydrology and Water Resources Management* (ed. by K. Kovar & H. P. Nachtnebel), IAHS Publication 211, 501-506.

Maidment, D R (1993) Developing a spatially distributed unit hydrograph by using GIS. *International Association of Scientific Hydrology Publications* 211, 181-192.

Monkton, C (1994) An investigation into the spatial structure of error in digital elevation data. In: *Innovations in GIS 1.* (ed. by M. Worboys). Taylor and Francis, 201-211.

Moore, I D (1993) Hydrologic Modelling and GIS. In:*Environmental Modelling and GIS II*, (ed. by M. F. Goodchild, B. Parks, C. Johnston & S. Glendinning), GIS World.

Moore, I D, Grayson, R B & Ladson, A R (1991) Digital terrain modelling: A review of hydrological, geomorphological and biological applications. *Hydrological Processes*, 5, 3-30.

Ordnance Survey (1995) Digital Data Catalogue. Ordnance Survey, Southampton.

Rieger, W (1993), Hydrological terrain features derived from a pyramid raster structure, In: *Application of Geographic Information Systems in Hydrology and Water Resources Management* (ed. by K Kovar and H P Nachtnebel), IAHS Publication 211, pp. 201 - 210.

Towers, W (1994) Towards a strategic approach to sewage sludge utilization agricultural land in Scotland. *Journal of Environmental Planning and Management*, 37, No 4, 447-460.

David R Miller
Macaulay Land Use research Institute
Craigiebuckler
Aberdeen
Scotland
AB9 2QJ
Tel: +44 1224 318611 x 2240
Fax: +44 1224 311556
Email: d.miller@mluri.sari.ac.uk

John C. Gallant and Michael F. Hutchinson

Towards an Understanding of Landscape Scale and Structure

ABSTRACT

Grid resolution of digital elevation data profoundly influences both the spatial pattern and the frequency distribution of derived topographic attributes, such as slope and specific catchment area, and therefore influences environmental models built on such attributes. A model of topography that explicitly incorporates scale is a prerequisite for an understanding of the origins and nature of the scale dependence of topographic attributes. We propose a model of surface topography that directly represents features in the landscape at a range of scales. This model allows generalisation and refinement of elevation models and also characterises the morphology - shapes, sizes and orientations - of features in the surface. Features are derived from a grid DEMusing a correlation detection process with successive refinement from broad scales to finer scales.

INTRODUCTION

Topographic analysis (or terrain analysis) is the quantitative analysis of topographic surfaces with the aim of studying surface and near-surface processes. A number of topographic attributes can be calculated, including specific catchment area, slope, aspect and plan curvature (contour curvature). Slope and specific catchment area are key variables in hydrology and are used to predict spatial patterns of soil water content and erosion (Beven and Kirkby, 1979; Moore et al., 1991; Moore and Wilson, 1992; Moore etal., 1993d; Moore, 1995; Wilson and Gallant, 1996). Solar radiation estimation is based on slope and aspect, modified by topographic shadowing (Moore et al., 1993c; Gallant and Wilson, 1996a). The spatial distribution of soil physical and chemical properties can be modelled within uniform geological settings using a combination of topographic attributes (Moore et al., 1993a; Gessler et al., 1995). Vegetation distribution, which responds to water, light and nutrient availability, can be modelled using combinations of topographic attributes which capture much of the landscape-scale variability of these parameters (Moore et al., 1993c). In short, topographic analysis provides the basis for a wide range of landscape-scale environmental models which are used to address both research and managementquestions.

It is now widely recognised that topographic analysis results are sensitive to the resolution of the source generalised. This affects all topographic attributes but in varying ways. The resolution-dependence of slope and specific catchment area have been the most intensively studied because of their regular application in hydrological modelling (Moore et al., 1993b; Zhang and Montgomery, 1994; Quinn et al., 1995). In these and similar studies the primary question to be addressed has been "What grid resolution should I use for a particular modelling exercise?", and some useful answers have emerged. But it is worth taking a step back and asking why topographic attributes are scale dependent and what that can tell us about the topographic surface. What are the characteristics of topographic surfaces that induce the observed scale dependence of topographic attributes? Is the scale dependence consistent across scales? Do topographic features have similar shapes at different scales? Is there a different kind of organisation of the landscape at different scales? In this paper we will explore some approaches to studying scale dependence of topographic surfaces.

SCALE ANALYSIS OF SURFACE TOPOGRAPHY

Our first observation is that grid resolution is not a particularly appropriate representation of scale. When we subsample an elevation grid to obtain another grid at coarser resolution, we are not only removing fine scale features of the surface (the intended change) but also changing the number of square cells into which the surface is divided. If grid resolution is used to study scale dependence of topographic attributes, the analysis is complicated by the different number of samples obtained from each resolution. Furthermore, specific catchment area is generally computed by accumulating cell areas from adjacent cells, and this network of connections is changed when the grid resolution ischanged. The minimum catchment area resolvable using the usual flow accumulation algorithms is also dependent on grid size, so grid resolution introduces a number of complicating artifacts to the analysis of scale dependence. If we wish to study the scale properties of a topographic surface, rather than the effect of grid resolution *per se* , it would be best to use a method that dealt with scale directly.

One technique for studying scale effects is spectral analysis, which provides information on relative amounts of variation at different wavelengths or spatial frequencies (Pike and Rozema, 1975; Gallant et al., 1994): wavelength is approximately equivalent to spatial scale. Spectral analysis as applied to topographic data is open to criticism on several fronts. Firstly, it assumes stationarity of the signal - the mean, variance and higher order moments should all be independent of location. This is clearly not true for topographic data, and in fact the spatial variation of such parameters is of considerable interest. The second difficulty with Fourier transforms is that they use oscillatory waveforms (sine functions) as the basis functions into which the sample data is decomposed. If the sine functions are not a good representation of the fundamental shapes occurring in the landscape then good localisation in scale is not possible. Non-sinusoidal shapes produce harmonics at shorter wavelengths which can overwhelm the contribution of smaller-amplitude features at those wavelengths. This also implies that a single feature in the landscape (if such a thing exists) might be represented in the Fourier transform by a substantial number of sinusoidal components, and conversely a single component in the Fourier transform contains contributions from a number of surface features. A basis function which better represents the fundamental shapes in the landscape would provide more meaningful representation of scale.

If we wish to overcome the first problem of non-stationarity we might use a small window which is moved across the data set to perform a spectral analysis at each window location. The main difficulty with this approach is that resolution must be traded off against location resolution, since a large window is needed to obtain good resolution in frequency (or wavelength) while a small window is needed to obtain good resolution in location. A more practical alternative is the wavelet transform which offers the best of both worlds by effectively using a window length that changes with wavelength. Short wavelengths use small windows thus giving good location resolution for fine-scale phenomena, while long wavelengths use large windows giving good wavelength resolution for broad-scale phenomena (Chui, 1992). Wavelet analysis uses a single basis function that is localised in both position and frequency and can translated and dilated to cover the entire position-frequency plane. This contrasts with Fourier analysis which uses many basis functions each at a different frequency. Wavelet analysis could produce a power spectrum at every point in the data set, which overcomes the non-stationary/non-local problem of the Fourier transform but results in a very large data set which then has to be processed further to obtain interpretable results. Furthermore, wavelet analysis still does not address the second problem of using oscillatory basis functions. Although there is a wide choice of basis functions, the wavelet transform requires the basis function to have zero mean which implies a degree of oscillation.

A MULTI-SCALE FEATURE-BASED REPRESENTATION OF TOPOGRAPHY

The representation we propose consists of a superposition of features at various scales. A surface can be constructed by introducing broad-scale features first and refining the surface by adding finer features onto the broader features. The features for such a representation should have at least the following properties:

- they must be simple enough that they have a recognisable intrinsic scale;
- they must be flexible enough that a variety of forms can be readily represented;
- the edges of the features must blend smoothly to zero so that continuity of the surface and its first derivatives is maintained.

The feature chosen to develop the method has an elliptical plan form and a smooth polynomial profile form as shown in Figure 1.

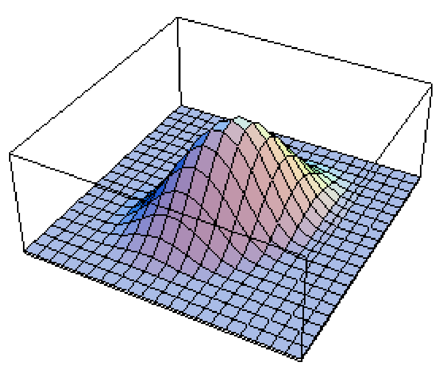


Figure 1. The elliptical feature chosen to represent topographic surfaces.

The parameters of a feature are its location, length, width, orientation and height: note that height can be negative. This is the simplest feature with sufficient flexibility to represent topographic surfaces. Figure 2 shows an idealised small catchment synthesised using 9 of these features to demonstrate that plausible topographic surfaces can be constructed using such features.

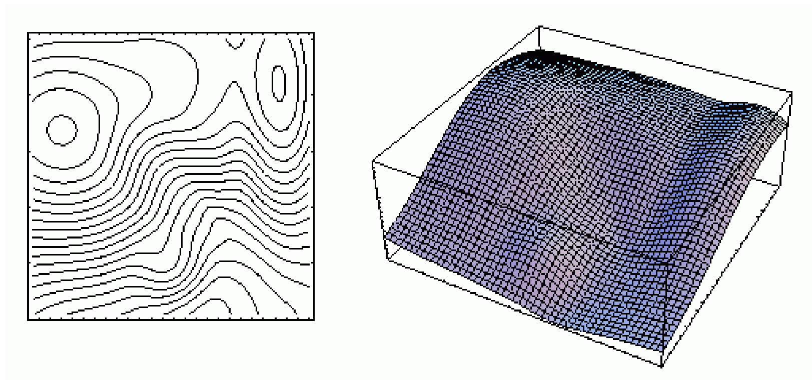


Figure 1. A small catchment synthesised using 9 features on a 1 km x 1 km square. Contours are at 20 m intervals.

Synthesis is the "forward" problem and is reasonably straightforward. How do we solve the "inverse" problem: is it possible to take an existing surface (typically represented as a grid DEM) and decompose it into a set of features that closely approximate the original surface?

DETECTION OF FEATURES USING POSITIVE WAVELETANALYSIS

Positive wavelet analysis (Watson and Jones, 1993) is similar to wavelet analysis in that it is based upon a single function which is translated and dilated, but it replaces the oscillatory wavelet with a positive pulse. Because this positive wavelet has a non-zero mean the inverse wavelet transform does not exist - it is not possible to reconstruct the original data from the wavelet coefficients computed from the forward transform. However, by using a process of correlation detection, individual features in the form of translated and dilated copies of the basis function can be detected in a surface which can then be used to reconstruct an approximation of the original surface. It is thus perfectly suited to the problem of decomposing surfaces using features.

The correlation detection algorithm finds the location, scale, shape and orientation of the feature giving the largest correlation value

$$T(x,y,L,w,\theta) = L^{-1}w^{-1/2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u,v)F(u-x,v-y,L,w,\theta)\,du\,dv$$

which is the real 2-dimensional wavelet transform of the function g(u, v) using the wavelet F(). The feature is located at x and y, L is its length, Lw is its width and theta is its orientation; note that the width w expresses the width of the ellipse as a fraction of length L and takes values between 0 and 1. The scale of a feature is L w^½ which is proportional to the square root of its area L^2 w pi/4.

The amplitude of a detected feature is

$$m = CL^{-1}w^{-1/2}T(x,y,L,w,\theta)$$

C is a constant dependent on the wavelet shape

$$C = \frac{1}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F^2(u, v, \cdot) du dv}$$

The detected feature is subtracted from the surface and another feature is detected from the residual surface. The parameters of features detected from the surface are distorted by the effect of other overlapping features, so the parameters of overlapping features are jointly optimised using non-linear least-squares (Marquardt, 1963; Press et al., 1989) with sum of squared differences between data and modelled elevations as the objective function.

The iterative decomposition detects the largest magnitude features first (which in topographic data tend to also be at large (i.e. broad) scales) followed by progressively smaller features. This leads to a relatively sparse representation of the surface being analysed provided the shape of the positive wavelet is a good match to the shape of the features in the sample. It is likely that some of the small scale features will be corrections for errors in the shapes of large scale components rather than real features in the data. If this is the case, some additional parameters could be included to modify the profile shape of the features to improve the fit to the surface.

In practice, scanning the entire 5-dimensional correlation surface T(x, y, L, w, theta) for every new feature is computationally very expensive. A simpler search has been developed which searches over a small subset of the space that is highly likely to contain the maximum correlation value.

RESULTS

Figure 3 shows contours of a 1 km x 1 km section from a 20 m DEM of the Brindabella Range near Canberra, Australia. The main features are the hill in the northwest corner, the general easterly slope on the eastern half and the drainage line running to the east in the southern half of the square.

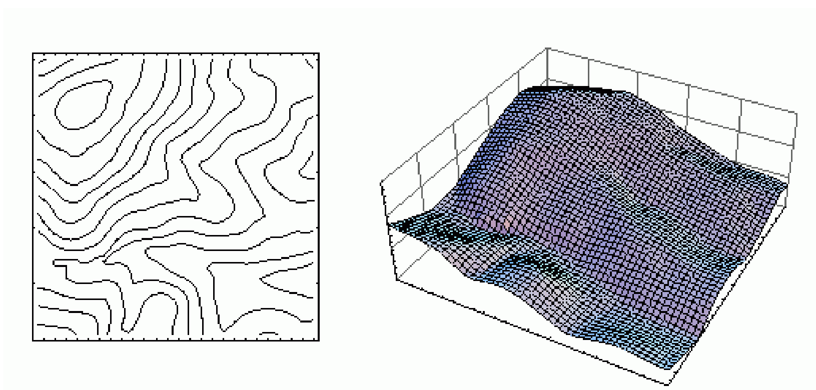


Figure 3. A 1 km x 1 km section from a 20 m DEM of the Brindabella Range near Canberra, Australia. Contour interval is 20 m.

Figure 4 shows a reconstruction of the surface using the first four features detected by the correlation detection algorithm. Note that these four features have captured the dominant broad-scale characteristics of this piece of landscape. Further refinement of the surface to 20 features is shown in Figure 5. The shape of the surface has been essentially captured but some fine scale detail is still missing. A decomposition to an average accuracy of 1 m (largest difference about 3 m) requires about 50 features for this patch.

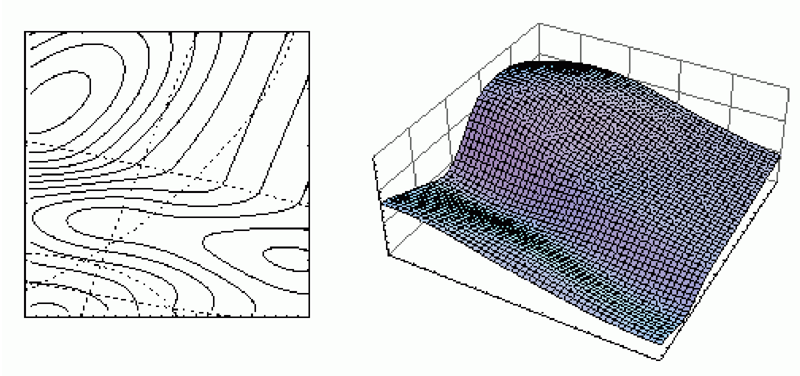


Figure 4. Reconstruction of 1 km x 1 km section using first 4 features detected. Contour interval is 20 m and dotted lines are outlines of 4 features.

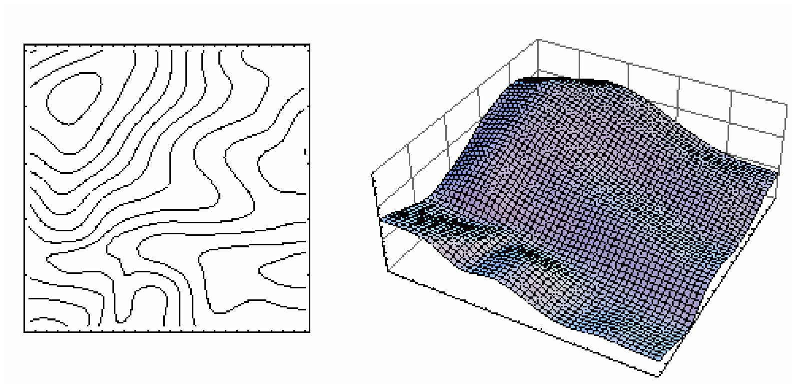


Figure 5. Reconstruction of 1 km x 1 km section using first 20 features detected. Contour interval is 20 m.

The patch of Figure 3 is too small to undertake a very meaningful scale analysis. The full 5.8 km x 3.7 km DEM from which the patch was taken has been decomposed to 690 features using a target average residual of 1 m. The surface was then reconstructed at four levels of generalisation by excluding features smaller than 200, 500, 1000 and 2000 m scale. All surfaces were reconstructed as 20 m resolution DEMs. The 2000 m scale surface contains only 12 features, the 1000 m surface has 37, the 500 m surface has 107 and the 200 m surface has 471 features. The contours of the original and the three most generalised surfaces are shown in Figure 6. The surface reconstructed using features larger than 200 m scale (not shown) differs from the original surface only at quite fine scales.

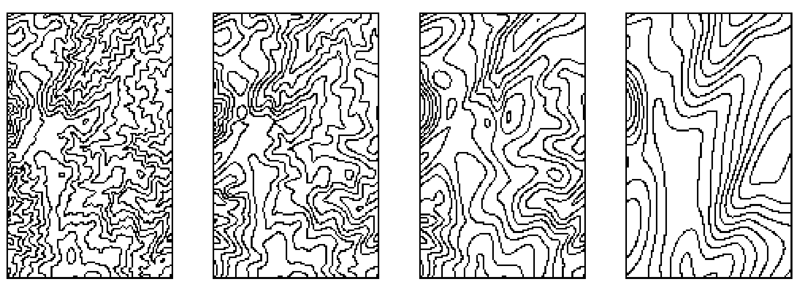


Figure 6. The 20 m resolution DEM of a 5.8 km x 3.7 km area in the Brindabella Range near Canberra, Australia. Contour interval is 50 m. From left to right: original DEM, and reconstructed surfaces with features larger than 500 m, 1000 m and 2000 m respectively.

The five surfaces were then analysed using the TAPES-G terrain analysis package (Moore, 1995; Moore et al., 1993b; Gallant and Wilson, 1996b) to compute contributing area using a multiple flow direction algorithm (Freeman, 1991; Quinn et al., 1991).The spatial pattern of contributing area from the original DEM and the 1000 m generalised surface are shown in Figure 7 with darker shades representing larger contributing areas.

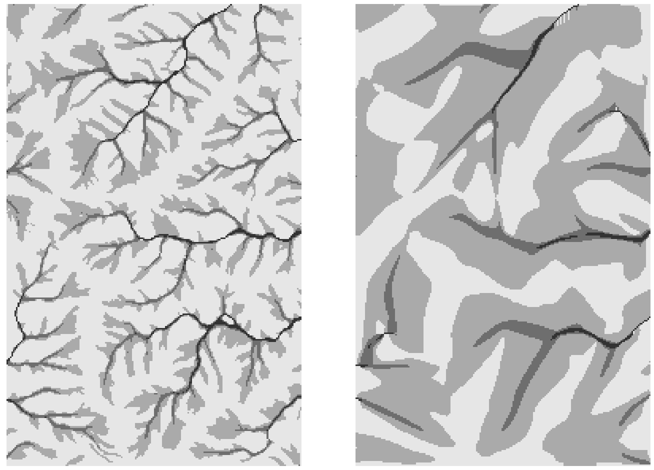


Figure 7. Contributing area computed from the original DEM (left) and from the surface reconstructed using features larger than 1000 m (right).

Figure 7 shows that much of the drainage network remains intact when features smaller than 1000 m scale (less than 1km^2 or 2500 grid cells) are removed, but that there is a much lower proportion of small contributing area values due to the absence of fine-scale division of the landscape into small catchments.

Figure 8 shows the frequency distributions for all 5 surfaces (original and four levels of generalisation) which also shows a progressive decrease of the proportion of small contributing area values as the surface becomes increasingly generalised. The most generalised surface shows a significantly different frequency distribution because it fails to separate two catchments on the eastern side of the DEM (see Figure 6. Previous analysis of the effect of scale on contributing area distributions using changes in grid resolution (Moore et al., 1993b; Zhang and Montgomery, 1994) have not been able to show the effect of scale at small contributing area because of the way minimum contributing area increases systematically with grid size.

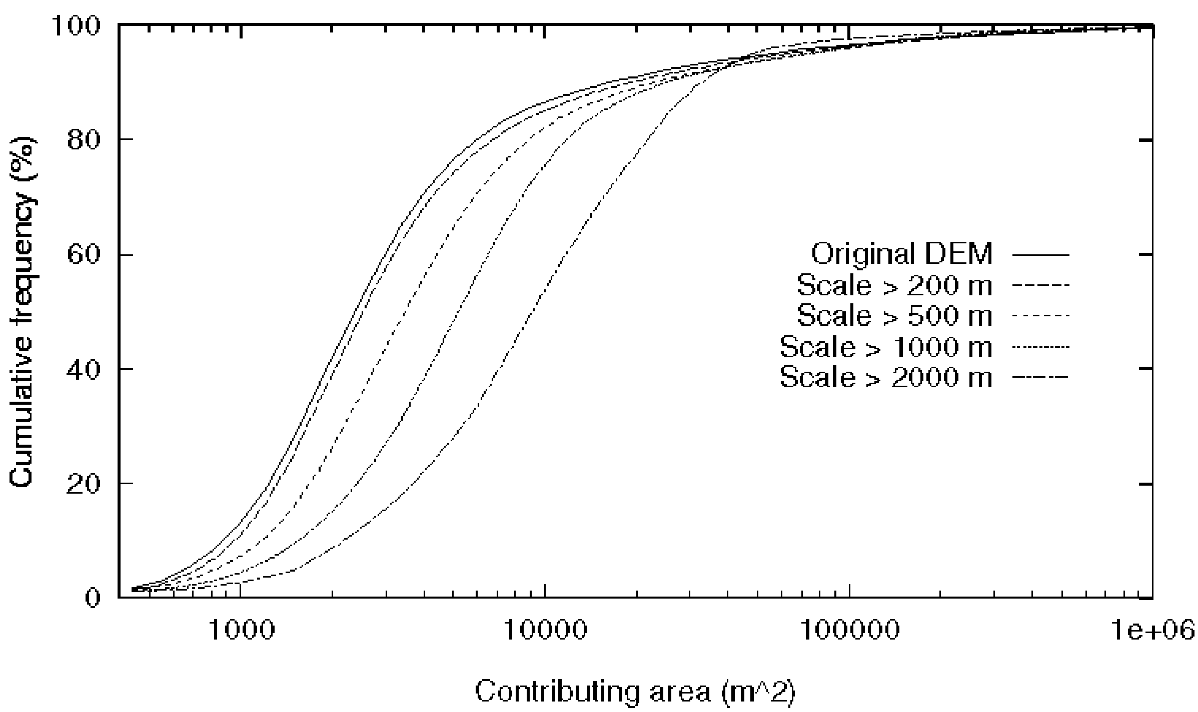


Figure 8. Frequency distribution of contributing area for original DEM and four generalised surfaces.

CONCLUSIONS

The positive wavelet decomposition presented here is a useful tool for analysis of scale dependence in topography. It explicitly identifies features at a range of scales, allowing generalisation of a topographic surface to allow detailed study of the effect of scale on topographic attributes without introducing artifacts due to changes in grid resolution. The shapes and orientations of features identified in the landscape may also be useful for characterising landforms and delineating regions of contrasting surface structure.

REFERENCES

Beven, K. and M. J. Kirkby (1979). A physically based variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.*, 24, 43-69.
Chui, C. K. (1992) *An Introduction to Wavelets.* Academic Press, Boston.
Freeman, T. G. (1991) Calculating catchment area with divergent flow based on a regular grid. *Computers and Geosciences,* 17(3), 413-422.
Gallant, J. C., I. D. Moore, M. F. Hutchinson, and P. E. Gessler. (1994) Estimating fractal dimension of profiles: A comparison of methods. *Math. Geol.,* 26(4), 455-481.
Gallant, J. C. and J. P. Wilson. (1996a) SRAD: A program for estimating radiation and temperature in complex terrain. *Computers and Geosciences,* (in preparation).
Gallant, J. C. and J. P. Wilson. (1996b) TAPES-G: A grid-based terrain analysis program for the environmental sciences. *Computers and Geosciences,* (in press).
Gessler, P. E., I. D. Moore, N.J. McKenzie, and P. J. Ryan. (1995) Soil-landscape modelling in southeastern Australia, in *GIS and Environmental Modeling: Progress and Research Issues,* edited by M. Goodchild et al. GIS World Books.
Marquardt, D. W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.,* 11(2), 431-441.
Moore, I. D. (1995) Hydrologic modelling and GIS, in *GIS and Environmental Modeling: Progress and Research Issues,* edited by M. Goodchild et al. GIS World Books.
Moore, I. D., P. E. Gessler, G.A. Nielsen, and G. A. Peterson. (1993a) Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.*, 57(2), 443-452.
Moore, I. D., R. B. Grayson, and A. R. Ladson. (1991) Digital terrain modelling: A review of hydrological, geomorphological and biological applications. *Hyd.Proc.*, 5, 3-30.
Moore, I. D., A. Lewis, and J. C. Gallant. (1993b) Terrain attributes: Estimation methods and scale effects, in *Modelling Change in Environmental Systems,* edited by A. J. Jakeman, M. B. Beck, and M. McAleer, chapter 8. John Wiley and Sons Ltd.
Moore, I. D., T. W. Norton, and J. E. Williams. (1993c) Modelling environmental heterogeneity in forested landscapes. *J. Hydrol,* 150, 717-747.
Moore, I. D., A. K. Turner, J. P. Wilson, S. K. Jenson, and L. E. Band. (1993d) GIS and land surface-subsurface process modelling, in *Environmental Modeling with GIS,* edited by M. Goodchild, B. Parks, and L. Stayaert. Oxford University Press.
Moore, I. D. and J. P. Wilson. (1992) Length-slope factors for the revised universal soil loss equation: Simplified method of estimation. *J. Soil and Water Cons.*, 475, 423-428.
Pike, R. J. and W. J. Rozema. (1975) Spectral analysis of landforms. *Ann. Assn. Amer. Geog.,* 82, 1079-1084.
Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. (1989) *Numerical Recipes. The Art of Scientific Computing.* Cambridge University Press, Cambridge. 702 pp.
Quinn, P., K.Beven, P. Chevallier, and O. Planchon. (1991) The prediction of hillslope flow paths for distributed hydrologic modelling using digital terrain models. *Hyd. Proc.,* 5, 59-79.
Quinn, P., K. Beven, and R. Lamb. (1995) The ln(a/tan beta) index: How to calculate it and how to use it within the TOPMODEL framework. *Hyd. Proc.*, 9, 161-182.
Watson, G. H. and J. G. Jones. (1993) Positive wavelet representation of fractal signals and images, in *Applications of Fractals and Chaos,* edited by A. J. Crilly, R. A. Earnshaw, and H. Jones, pp. 117-135. Springer-Verlag, Berlin.
Wilson, J. P. and J. C. Gallant. (1996) EROS: A program for estimating spatially distributed erosion indices. *Computers and Geosciences,* (in press).
Zhang, W. and D. R. Montgomery. (1994) Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resour. Res,* 30(4), 1019-1028.

John Gallant johng@cres.anu.edu.au and Michael Hutchinson hutch@cres.anu.edu.au
Centre for Resource and Environmental Studies
Institute of Advanced Studies
Australian National University

Canberra ACT 0200
Australia
Fax: +61 6 249 0757

Dean Djokic, Andrew Coates, and James E. Ball

GENERIC DATA EXCHANGE - INTEGRATING MODELS AND DATA PROVIDERS

The integration of GIS with modeling software has usually been a one-off exercise, with large amounts of time spent creating and maintaining specialized interfaces and data exchange formats. This paper presents a generic format and method for data exchange (GOODES), which can significantly streamline the integration process between information users and information providers. Information users are usually analysis or modeling packages, such as hydrologic models or statistical packages, but can also include GIS when used as an analysis tool. Information providers are usually data storage and retrieval systems, but can also include models when used to feed back information (results) from the simulation. GOODES stands for generic, object-oriented, open data exchange system, and allows exchange of data between any platform or package through an open and object-oriented structure. Application-specific drivers are used to translate the data to and from the generic format and can be developed for any application as required. Examples of data exchange between GOODES format and ARC/INFO GIS, hydrologic/hydraulic model SWMM, and time-series database HYDSYS are provided to illustrate the system's implementation and features.

Introduction

Integration of computer models describing various processes occurring in the environment is becoming a major obstacle in development of comprehensive planning and analyses tools that can handle complex process interactions. Most models that have been developed over the last 30 or so years have concentrated on solutions of small portions of the overall problem in a field of engineering.

For example, in hydrology, the models have often been developed for a single component of a hydrological cycle, such as groundwater flow, surface flow, or flow in unsaturated zone. In many cases even these partial solutions are limited in their scope, often based on the dimensionality of the problem (1-D, 2-D, 3-D, time variant or invariant). There have been few attempts to create comprehensive solutions to general problems (e.g. the whole hydrological cycle), and those solutions usually compromise certain aspects of detail of the overall problem.

The main reasons for such approach are:

- complexity of the problem, requiring simplifications due to our lack of knowledge about the process.
- complexity of the solution, requiring too much computer resources for a comprehensive solution.
- lack of data to support a complex solution.

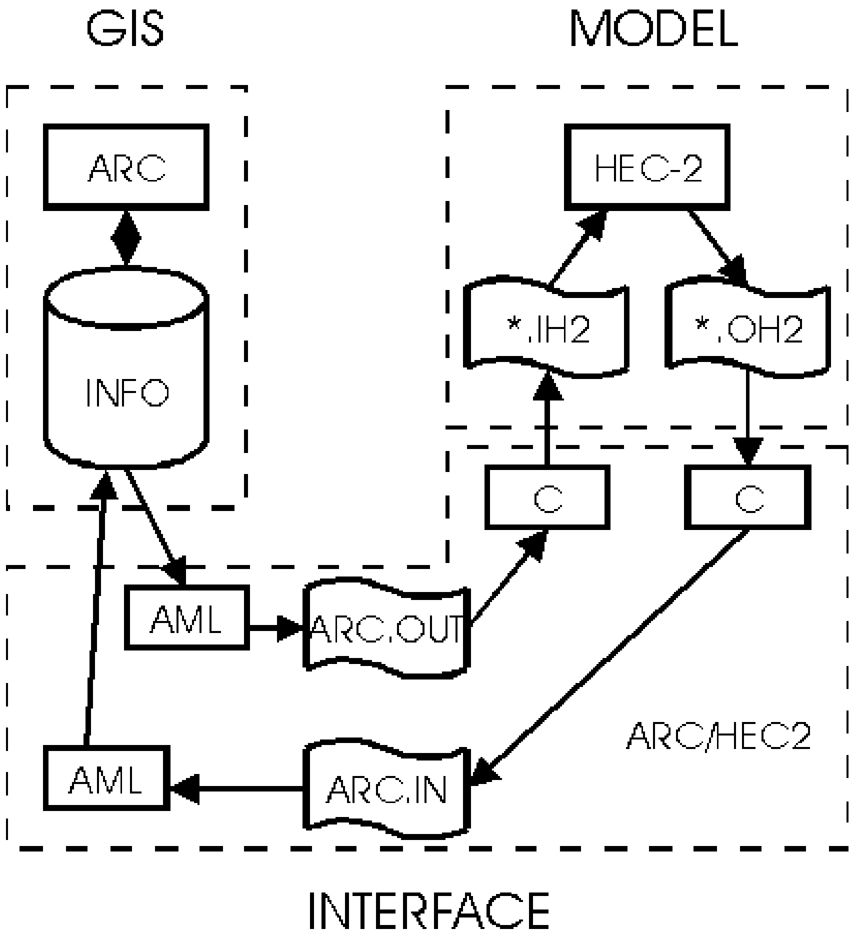
In last few years, with wider use of remote sensing as data provider for model input and fast development of computing power, two of the three indicated reasons are slowly being removed, allowing analyses of more complex problems. Those interested in such analyses are faced with the dilemma on the approach to take in developing complex models. There are two main options:

- start from the beginning and rewrite all the pertinent code from the ground up. This approach will produce the most efficient code but could require an unreasonable or unavailable amount of time for completion. Other aspects, such as legal issues, acceptance in the user's community, lack of expertise, requirements for retraining and others, could preclude use of this approach.
- use existing programs that already solve individual components of the problem and make them work with each other. The benefits of this approach are that all the existing efforts in developing the solutions can be reused, there is (hopefully) acceptance of the model in the user's community, they have been thoroughly tested (possibly even in court), etc. The major drawback is the need to work with each individual model and understand its internal structure and operation.

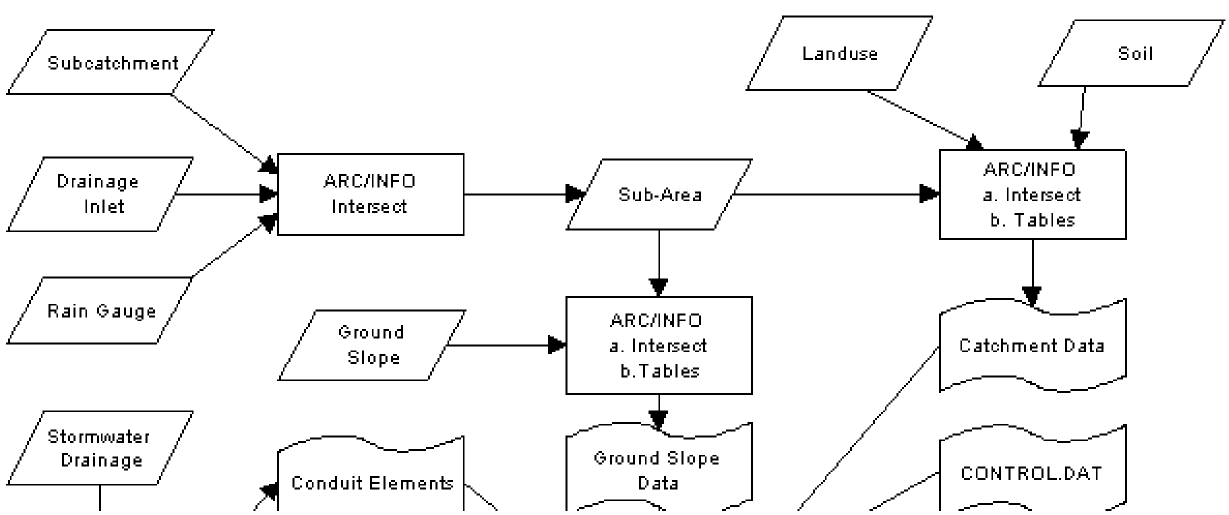
When analyzing the requirements for development of modern problem solving tools in environmental engineering, the second approach seems more rational (Djokic, 1993). An opportunity to use the first approach will arise as the need for new solutions to the old problems comes about due to better understanding of the basic underlying principles. In some areas of our endeavor that is already happening, but in some others, the existing techniques will be used in their present form for a long time to come.

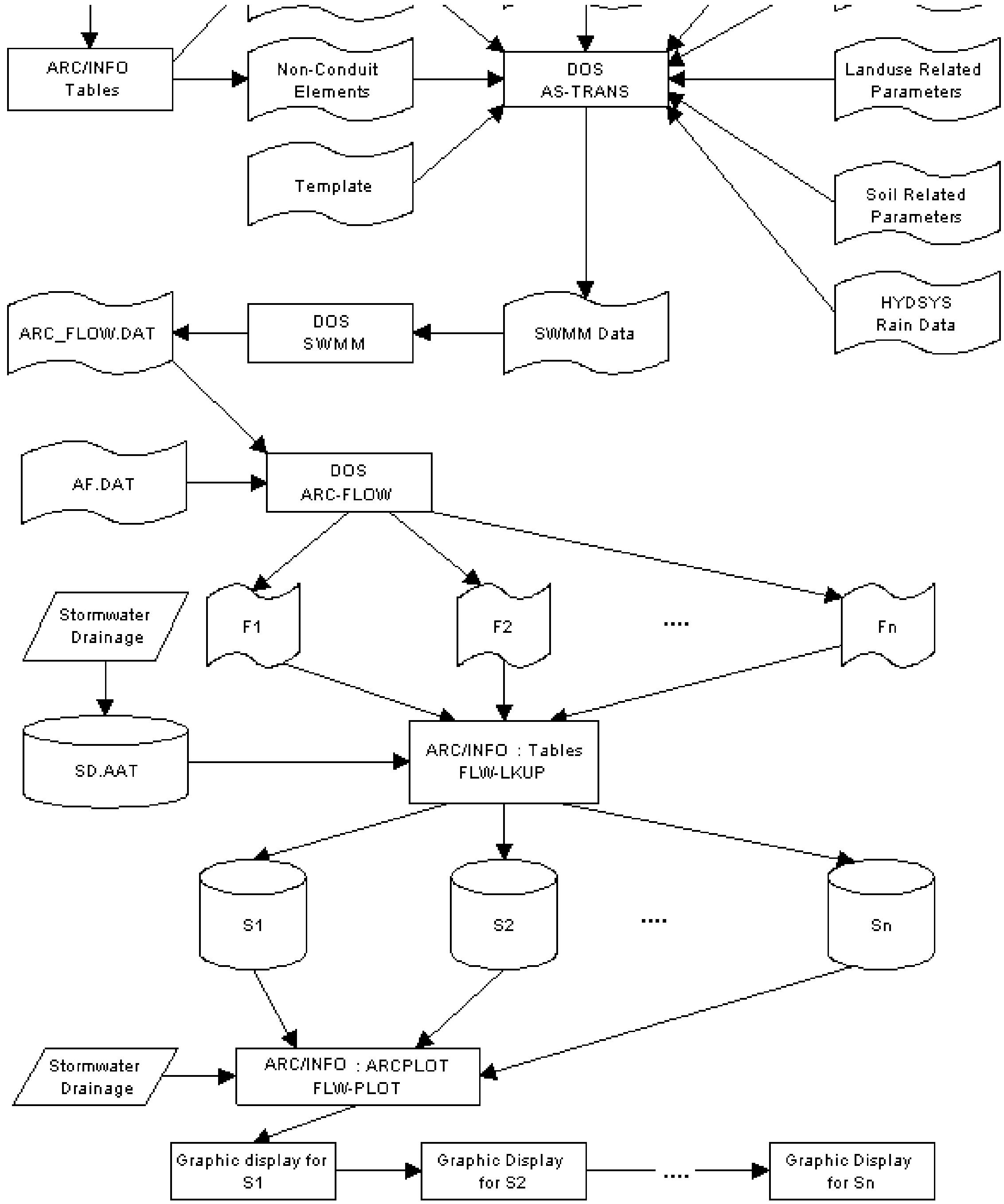
Need for Generic Data Exchange Format

A complex decision making tool can consist of several components that are more or less tightly integrated. Depending on the accessibility of the source code for the programs of interest and interface strategy, the two programs can be linked, integrated, or embedded (Djokic et al., 1995a). Integrating computer programs on a data level is a tedious, but not very difficult task.



An example of an application in which ARC/INFO GIS and HEC-2 hydraulic model are integrated using several interface programs and temporary data transfer files (Djokic et al., 1994) is provided in figure 1. The figure is a schematic representation of the interface. Actual implementation consists of more than 30 different programs, macros, and temporary files. As the number of programs to be interfaced increases, the complexity of the data exchange system increases significantly. Figure 2 presents a detailed data exchange structure between ARC/INFO GIS, HYDSYS temporal database, and SWMM hydraulic model (Chui, 1995).





Standard practice is that for each pair of programs to be integrated, a separate, program specific, data interface is created. Although convenient for development, such approach makes maintenance of complex interfaces difficult, especially as individual components (programs) go through version changes. Rationalization of the approach can be made by using standard data transfer file formats (Djokic et al., 1995a). Ability to use these standards can greatly reduce required effort in data integration, but is often limiting in the type of the data that can be transferred.

There are several standards available for data exchange between different computer programs and program/data types. Most of them are program/data type specific, such as SDTS (USFIPS, 1992) for exchange of spatially distributed data, DXF (Thomas, 1989) for CAD type of data, HDF (NCSA, 1995) for raster based data, or product developer specific.

The main problem in use of these standards is their inflexibility to handle diversity of data types encountered in environmental models. For example, depending on the problem being analyzed, an apparently same measure will be treated differently by different models. Consider a case of conjunctive surface water, groundwater use. The watershed boundaries, possibly delineated by hand, can be stored in a GIS. Surface water and groundwater boundaries can be different and stored as different coverages. A modeling program can retrieve these areas and use them for detailed hydrologic/hydraulic computations. As a result of the model calibration, it is possible that the initial watershed areas have been changed. We now have several places that hold a (different) number representing the same spatial feature.

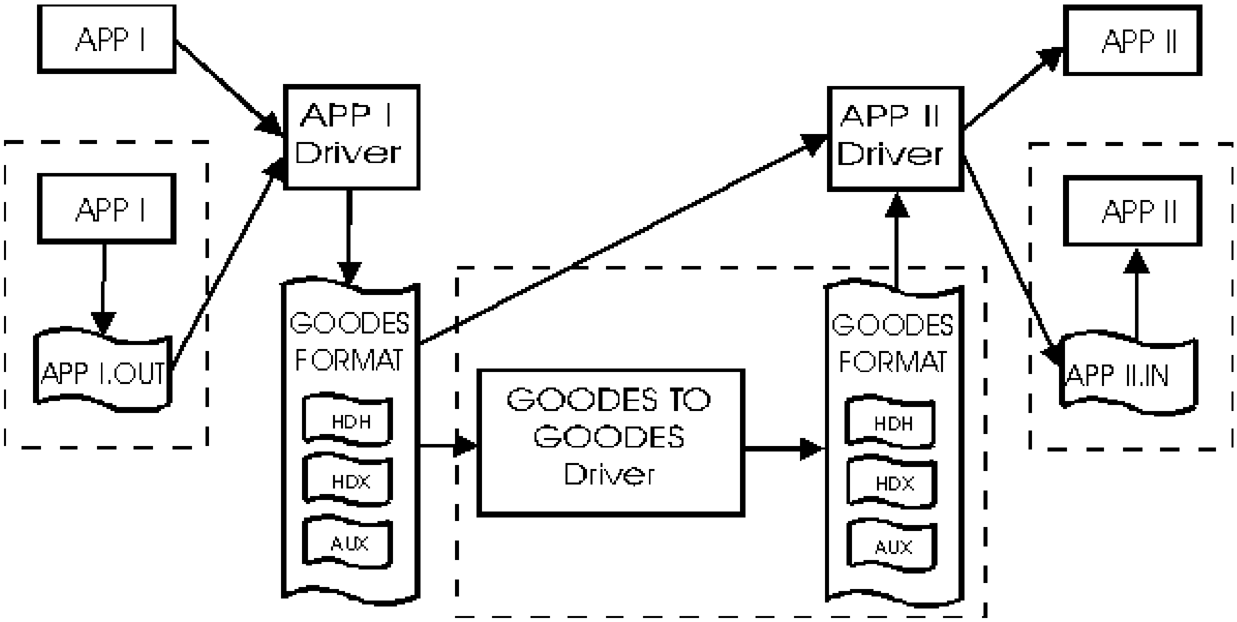
Now, another application wants to use the watershed areas for further analyses. Several issues arise:

- where will that application look for the relevant data: in GIS coverages, or in hydraulic/hydrologic model database?
- which data should be trusted: the calibrated ones, the estimated ones, or a combination?
- should the problem of having different values for the same entity be somehow resolved, and if so, how?
- if a standard data exchange format is used, is it possible to attach some intelligence to the transferred data, so that the new application can react to the origin of the data - maybe treat the datum as a deterministic value if determined by one method, or as a fuzzy value if determined by another method?

Such questions make use of existing data exchange standards difficult. This does not mean that the existing standards should not be used. They certainly should. It is important however to realize that for a number of cases they will not be sufficiently flexible.

Exchange Format Definition - GOODES

To accommodate the diversity of data and models that could be integrated a new data exchange method has been proposed (Djokic et al., 1995a). To acknowledge the generality of the method and its structure it has been renamed to GOODES - generic, object-oriented, open data exchange system. GOODES consists of two major components. First is a data exchange file structure, described in detail by Djokic et al., 1995a, with more details published on WWW at http://www.water.civeng.unsw.edu.au/department/hydrology/goodes2.htm. This structure consists of up to three ASCII files that define data structure and character, actual data, and all necessary auxiliary information needed for data interpretation.



The second component of GOODES method are the drivers that actually convert GOODES files into application understandable format, or create GOODES files from the application. Figure 3 is a schematic representation of GOODES exchange system. The role of the drivers is more than just reformatting the data, and it will be discussed in detail in the following section.

GOODES Drivers

There are three distinct GOODES driver types. These fulfill all of the requirements for data exchange and are intended to simplify the process of application integration. The driver types are:

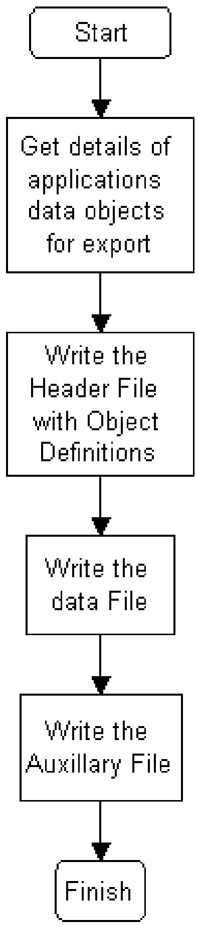
- Application to GOODES
- GOODES to Application
- GOODES to GOODES

Application to GOODES

The application to GOODES driver simply puts data objects from the application into GOODES format. The driver is generally written in the application's own language if there is one available, as this will usually simplify access to the native data objects. The application to GOODES driver will generally work from the application's internal data structure if that is exposed to the author, however for some applications, there may be no access to that structure and the driver will have to work from standard output from the application. There is no processing performed on the data objects (other than to specify which are to be exported) as the driver assumes no knowledge of the target application.

The sequence for exporting the application's data to GOODES format is shown in Figure 4. The user first supplies details of the application's data objects required for export. This specification can either be via a user interface or a controlling script file. Next, the user provides any additional information required for the class definitions in the header file, which is then written.

The global section details are provided and the data file written. Finally any information for an auxiliary file is written.

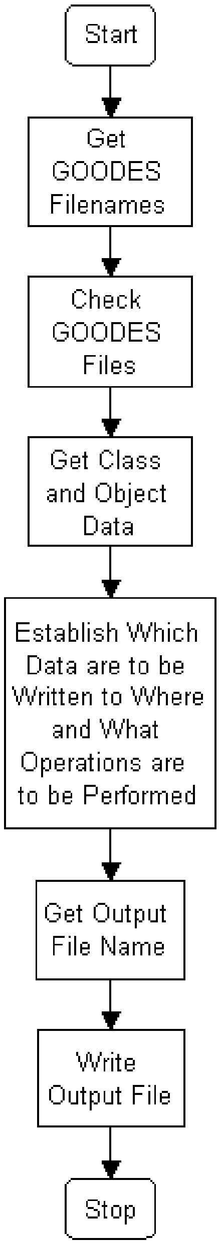


GOODES to Application

The second driver type takes information in GOODES format and converts it to the application's input data format. This may be the application's native format if that is available to the author, or may be an intermediate format with which the application is familiar. By choice, the GOODES to application drive will be written in the native language of the application if there is one, as this will generally provide the easiest access to the applications data structure.

The GOODES to application driver will generally have some limited amount of 'intelligence'. It should, for example, be able to carry out basic database operations such as single level joins, aggregation, summary, some simple statistical analysis and boolean selection. To achieve this, it will often be required to go through an intermediate data format (such as import into a RDBMS) if the application itself does not support such operations. This intermediate format is not intended to be a standard and its structure is entirely at the discretion of the driver author.

Generally, much more user interaction is required for the GOODES to application driver than for the application to GOODES driver. This suggests that a facility for script file processing of data exchange, as well as manual interaction be provided.



The process for the GOODES to application driver is shown in Figure 5. First, the user supplies the name of the data file (either interactively or through the controlling script file) and the driver checks that it, the header and any auxiliary files are available. The structures of the files are then checked for consistency. The performed checks are listed in Table 1.

- Data File
  - Global section contains header file name and date format.
  - All data objects have corresponding class definitions in the header file.
  - Data objects have unique names.
- Header File
  - All class definitions are in the correct format.

- Classes have unique names.

GOODES to GOODES

The final driver type is used for complex data manipulation tasks. It operates on one or more GOODES files and either produces a new GOODES file or appends its results to an input file. The GOODES to GOODES driver is used when the format of the origin data cannot be resolved into an appropriate structure by the simple operations available through the GOODES to application driver.

The GOODES to GOODES driver supports complex relational database operations, and as such will usually use an intermediate data format such as importing the data objects into tables in a RDBMS. This format is entirely at the discretion of the driver author and is not in itself a standard. The driver should be able to be controlled either interactively or through the use of a script file. Actions performed interactively should be recordable in the script file format for later automatic repetition.

Some of the database operations which could be supported by a GOODES to GOODES driver are :

- Multi-table relational joins
- Advanced mathematical and statistical functions
- Data conversion
- Spatial transformation or projections
- Complex boolean selection based on fields in multiple classes

Case Study

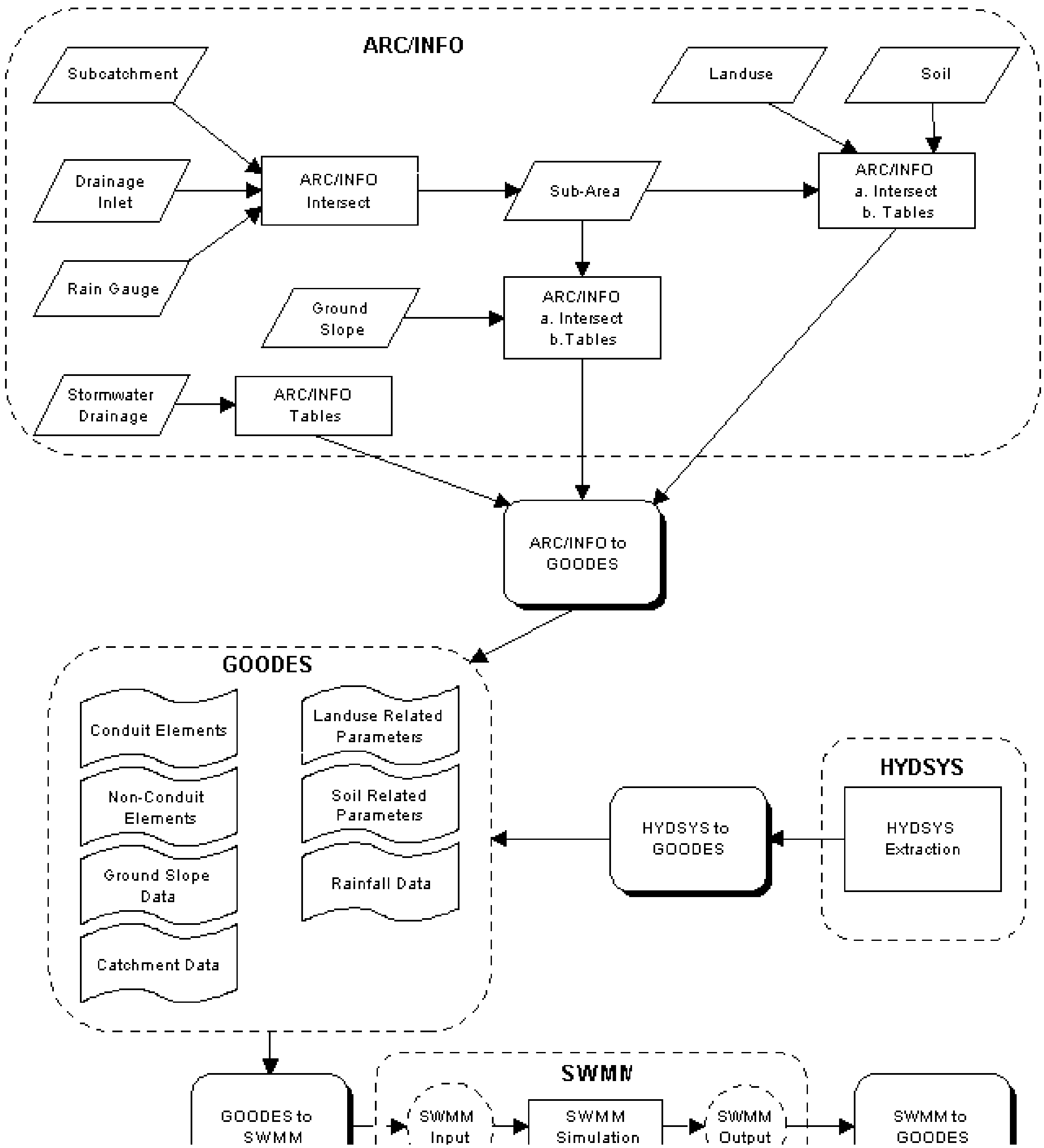
An existing one-off integration exercise was used as a case study for this generic approach to model integration. This exercise, which is described in more detail in Djokic et al., (1995b) and Chui (1995), involved the use of data from ARC/INFO about the spatial elements in an urban catchment and from a time series manager, HYDSYS that held the information about rainfall. The data were combined to run simulations in the EPA's Storm Water Management Model (SWMM), and to feed the results back into ARC/INFO for display. This was achieved through some custom written integration routines in AML and QBASIC. A schematic diagram of the integration process is shown in Figure 2. As can be seen, several custom-written interface programs were written. The case study involved the replacement of these with new components based on the GOODES format. In particular, the following drivers were written:

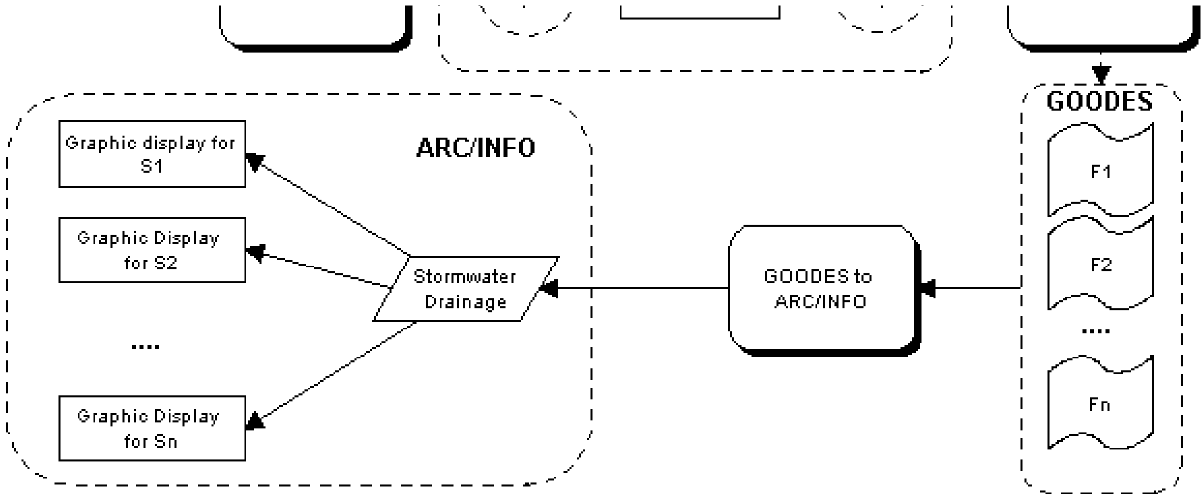
- GOODES to Application Drivers
    - GOODES to ARC/INFO
    - GOODES to SWMM
    - GOODES to FoxPro
- Applications to GOODES Drivers
    - SWMM to GOODES
    - ARC/INFO to GOODES
    - HYDSYS to GOODES

A schematic of the resultant process after the generic method was applied is shown in Figure 6.

As can be seen from a comparison of the two figures, the generic method produces significantly enhanced results in terms of clarity of the process.

As a result of this generic approach, several benefits have ensued. Firstly, each of the components is now modularised. Should the requirements of the system change such that, for example, a different time series manager would be preferable, only that small part of the diagram relating to the HYDSYS components need to be re-written. In fact, with the generic approach, it is likely that the driver for the new time series manager is already available (perhaps having been developed for some completely unrelated project) and the process of replacing components is extremely simple.





The driver development for the case study also clarified the philosophy behind the three driver types (and actually revealed the need for a GOODES to GOODES driver). At first, the application to GOODES drivers were perceived to require some knowledge of the target application in their object structure. While it is still true that a very general knowledge of the purpose of the output data is required, it was decided that a truly generic system would not impose structures on export. The import drivers should manipulate the GOODES data, rather than relying on it being pre-formatted. To reduce the need for extremely complex capabilities for all GOODES to application drivers, the GOODES to GOODES driver was developed. It is a common module that can be used prior to GOODES to application drivers, providing the manipulative, summary, and other capabilities that may be required, without the need for complex development effort by all (GOODES to application) driver authors.

Discussion and Conclusion

GOODES data exchange system facilitates data integration of applications with diverse background. GOODES file type and structure enables easy cross-platform and cross-application access to the data presented for sharing. Additional information about the data helps in proper interpretation of the meaning of the data, which is crucial for development of complex, integrated modeling systems that require little or no user intervention.

Drivers add intelligence to the data exchange process. They are used to interpret the data as a function of the data character and auxiliary information stored in GOODES transfer files. It is envisioned that drivers to and from an application will be developed and maintained by the application developers. Since it is impossible for any developer to envision and accommodate all the possible uses of their product, often the GOODES files will not be complete enough for direct transfer to another application.

A GOODES to GOODES driver, possibly in conjunction with GOODES files derived from some other programs, can be used to create a GOODES file that an GOODES to application driver can interpret directly. This approach allows the application integrator to concentrate on the major issues of the data integration process. That involves fine tuning of the GOODES to GOODES driver for specific problem and programs at hand, but as a benefit, it relieves the application integrator from dealing with details of individual programs' input and output structure. As the number of programs to be integrated increases, this benefit becomes more pronounced.

The GOODES system is public and open for users' input. It is documented and will be maintained through WWW entries at http://www.water.civeng.unsw.edu.au/department/ hydrology/ goodes2.htm. Comments and submissions are welcome and appreciated. The contact regarding GOODES maintenance is Andrew Coates (A.Coates@unsw.edu.au)

References

Chui, S.K., (1995). Hydrologic Application of ARC/INFO GIS to SWMM for Modelling Urban Stormwater Problem. M.Eng.Sci thesis. School of Civil Engineering, University of New South Wales, Sydney, Australia.

Djokic, D., Coates, A., and Ball, J.E. (1995a). "GIS as Integration Tool for Hydrological Modeling: A Need for Generic Hydrologic Data Exchange Format." 15th Annual ARC/INFO User Conference, Palm Springs, CA.

Djokic, D., Ball, J.E., and Chui, S.K. (1995b). Integration of ARC/INFO and Stormwater Management Model (SWMM). 9th Annual Australian ARC/INFO User Conference, Sydney, Australia.

Djokic, D., Beavers, M.A., and Deshakulakarni, C.K. (1994). "ARC/HEC2: an ARC/INFO - HEC-2 Interface." Proc. 21th Water Resources Planning and Management Division Annual Specialty Conference, ASCE, New York, N.Y., 41-44.

Djokic, D. (1993). "Towards General Purpose Spatial Decision Support System Using Existing Technologies." NCGIA Second International Conference/Workshop on Integrating GIS and Environmental Modeling, Brackenridge, CO.

NCSA. (1995). Hierarchical Data Format. National Centre for Super Computing Applications, WWW, http://www.ncsa.uiuc.edu/SDG/Software/HDF/HDFIntro.html, Urbana-Champaign, Il.

Thomas, R.M. (1989). Autocad Desktop Companion. Sybex, Alameda, Ca.

USFIPS. (1992). Spatial Data Transfer Standard. US Federal Information Processing Standard Publication 173.

Dean Djokic, Lecturer
Andrew Coates, Professional Officer
James E. Ball, Senior Lecturer
School of Civil Engineering
The University of New South Wales
Sydney, NSW 2052
Australia
Telephone: 61-2-385-5771
Fax: 61-2-385-6139
Email: D.Djokic@unsw.edu.au

Stephanie L. Greene, Thomas Hart

Plant genetic resource collections: an opportunity for the evolution of global data sets

Abstract

Decline in global biodiversity threatens plant genetic diversity, the raw materials we rely on for food, fiber, medicine and industrial products. Complementary conservation strategies couple the protection of wild plant populations and traditional crop varieties where they have evolved, with the collection and preservation of genetic diversity in gene banks. As part of the Agricultural Research Service of the USDA, the National Plant Germplasm System (NPGS) is responsible for a genetic resource collection containing 444,724 accessions. Germplasm is maintained in the form of seed or live plants, representing current, obsolete and primitive crop varieties, wild and weedy relatives of crop species, and wild species collected from around the world. The NPGS has developed the Germplasm Resources Information Network database (GRIN), that describes collection holdings. For a given accession, information in GRIN can include ecogeographic description and location of collection site; taxonomic, morphologic and historic cultural characterization; response to disease, insect and environmental stress; agronomic suitability and availability of seed. Efforts are underway to improve the quality and quantity of information within GRIN. Currently, GIS use is constrained by limited georeferenced information. A diverse group of curators are working to include historic locality information in GRIN, but the need remains to develop a straightforward approach to estimate collection site map coordinates based on distance and direction from cited settlements. Using GIS technology, curators and users can develop external purpose-built programs to link collection information to a broad array of spatial information. Spatial analysis of collections can lead to a greater understanding of the ecogeographic representation of existing collections, increasing the efficiency of conserving and managing genetic resources and broadening our understanding of the distribution of genetic diversity. Genetic resource databases could potentially serve as an information resource to a broader scientific community, providing relatively independent global data sets to meet the unique needs of environmental modelers. Multi-disciplinary collaboration would ensure the evolution of a set of databases useful to a broad range of professions.

Introduction

Biotic diversity provides the raw materials we rely upon for food, fiber, medicine and industrial products. Plant germplasm, the inter- and intraspecific reservoir of potentially useful genetic materials, is an essential natural resource that provides insurance against crop pests and environmental stresses that impact agricultural production. Declines in global biodiversity threaten plant diversity at the species level, and within species, at the genetic level. Complementary conservation strategies include the protection of wild species, plant populations and traditional crop varieties where they have evolved (*in situ* conservation), with the collection and preservation of inter- and intraspecific diversity in gene banks and botanical gardens (*ex situ* conservation). *Ex situ* genetic resource collections maintain germplasm in the form of seed or live plants, representing current, obsolete and primitive crop varieties, wild and weedy relatives of crop species, and wild species collected or donated from around the world. This material is conserved, but is also available to a broad scientific community for basic research and development into crop cultivars.

On a global scale, Williams (1989), estimates there are 400 significant plant genetic resource collections. An estimated 1,787,024 accessions are held at the five largest national collections and at the major International Agricultural Research Centers (NRC 1993). As part of the Agricultural Research Service of the USDA, the National Plant Germplasm System (NPGS) is responsible for a genetic resource collection containing 444,724 accessions. The formation of the NPGS can be traced back to 1898, with the establishment of the USDA Section of Seed and Plant Introduction (White et al. 1989) . Historically, the NPGS and other national programs focused exclusively on the collection, evaluation and introduction of crop plants having economic potential. The Green Revolution in the 1960's which resulted in the development and widespread acceptance of high yielding hybrids, was the initial impetus in shifting the focus of germplasm activities from "exploitation" to "conservation" in an effort to preserve traditional varieties from displacement (Cohen et al. 1991). The 1970 U.S. epidemic of corn leaf blight served as a catalyst for shifting the NPGS focus to the conservation and utilization of genetic resources as a guard against genetic erosion and genetic vulnerability. The current NPGS was organized in the early 1970's and is characterized as a network of germplasm repositories, consisting of about 24 sites which actively store, grow out and distribute germplasm collections. The National Seed Storage Laboratory in Fort Collins provides safety back up for the collections.

Information associated with germplasm accessions is as important as the actual plant material held in gene banks (Given 1994). Collection documentation facilitates not only the management but use of germplasm collections. To varying degrees, most genetic resource collections have developed and continue to improve databases that describe collection holdings. The objectives of this paper are twofold: 1.) to provide a detailed description of plant genetic resource databases, using as an example the database that has been developed by the NPGS; 2.) to discuss the implications of linking germplasm collection information to a broader array of spatial data from two perspectives: the genetic resources community, and a broader community embracing a range of professions.

Germplasm Resources Information Network

Overview

The Germplasm Resources Information Network (GRIN) is the centralized database used by the U.S. National Plant Germplasm System to house information on individual collections and support system-wide activities. GRIN is set up on a SUN microcomputer set up with 45 gigabytes of disc space and two 135 MIPS processing units performing under a UNIX operating system. The Oracle relational database has been used to develop GRIN. Public users can access GRIN through a modem, or through the internet via the World Wide Web or gopher servers. A crop specific PC-based version of GRIN, containing collection information and a stand-alone menu-driven software package can be downloaded from the internet, or obtained as floppy disks. Detailed information for accessing GRIN can be found in Appendix 1.

The GRIN database has been organized into 8 general areas. Central to the database is the ACCESSION (ACC) area. Each accession is assigned a unique identifier (PI number) and appears as a single record. The ACC area contains passport data, information which describes where and when an accession was collected, donated or developed, and the individuals responsible for the acquisition. For field-collected accessions, a description and location of the collection site is may be available. Location of collection sites are most frequently cited in terms of distance and direction from the nearest settlement. Collector comments such as number of plants collected or local name and use of germplasm may also be included. For donated and developed accessions, information is frequently available that describes the genetic background of an accession, alternative names (ie. other institute identifiers), and status of intellectual property rights. In addition, historic taxonomy changes, quarantine status and literature citations also appear in the ACC area. The OBSERVATION (OBS) area contains descriptor data for each accession. Crop-specific descriptor lists have been developed for most crops to provide a means of comparing accessions within a collection based upon standardized morphological, phenological or physiological traits, as well disease and insect stress responses. Common descriptors include flower color, leaf size, growth habit and plant height and width. The EVALUATION (EVAL) area contains data on more detailed studies, providing a description of the study environment. scienticists involved and literature citations. Seed and plant material coming into and out of the system are tracked in the INVENTORY (INV) area. Details regarding the status of specific seed lots or plant samples for a given accession, such as current quantity on-hand, viability and availability are monitored from this area. Seed and plant requests are processed and tracked from the ORDER PROCESSING (ORD) area. Each accession is linked to the TAXONOMY (TAX) area, which provides detailed taxonomic information for each accession. All areas are linked to the COOPERATOR (COOP) area, which maintains the names and addresses of persons involved with the system, and the STANDARDS (STAN) area, which maintains standardized tables for information such as geography, literature, and crop names.

Data limitations

The quality and quantity of passport and evaluation information in GRIN varies widely among crop collections, and within collections, among accessions. Since the early 1980s, data from evaluation studies has been routinely placed in GRIN. Users can query the database to identify accessions within a crop collection with attributes of interest. The amount of evaluation data varies considerabley depending upon the trait evaluated, and on the specific species examined (Greene and Bohning 1994). The user should be aware that frequently only subsets of collections have been evaluated for a given trait. This is in part due to the unavailability of accessions with limited seed quantities, and prohibitive costs of evaluating a large number of accessions. Generally, the major cultivated crop species have been much more extensively evaluated than non cultivated species.

Major crop species generally have more complete passport information as well. Recent acquisitions (since the mid-1980s) have the most comprehensive documentation. As new acquisitions are received, passport data is entered into GRIN as a routine part of processing. Unfortunately, for many accessions acquired prior to the development of GRIN, computerized documentation is limited. Germplasm that has been donated to the NPGS generally has much less information than germplasm collected in the field (Greene and West-Smith 1994). Collection site information, including latitude and longitude values, and site locality and habitat description is particularly relevant for using GIS. Table 1. provides a summary of collection site information available in GRIN for 6 major crops. Most limiting is the availability of latitude and longitude values. Currently, only 9 % of the accessions cited in Table 1 have this information in GRIN, although there is considerable variation among collections for the amount of data available. For example: 17% of sorghum accessions, but only .3 % of the rice accessions currently have this information in GRIN. Since the early 1990's, global positioning systems (GPS) have been routinely provided for NPGS-funded trips, so recent acquisitions usually have latitude and longitude values.

An average of 50 % of the accessions cited in Table 1. have a narrative description of the collection site location. The distribution of locality data varies among the collections. For example, 72 % the sorghum collection has locality information, as oppposed to 9 % of rice accessions. Efforts have been made to estimate site coordinates using an atlas reference, based on the settlement cited in passport locality information (Steiner and Greene 1996, Steiner and Poklemba 1994, Pederson et al.1996). Problems encountered included lack of settlement references (because of settlement name changes, small size or collector inaccuracy) and inaccuracy due to significant distance ofcollection site from cited settlement. Such procedures are labor intensive, making it difficult to apply to large collections (Steiner and Greene 1996). The development of software that could automate the conversion of passport locality information into collection site map coordinates based on distance and direction from cited settlement, would be valuable. To ensure data integrity, estimated point localities would need to be verified and documented as such in GRIN (Steiner and Greene 1996). On-site ecogeographic description is also limited and varies with collection. For example: 20 % of the accessions in Table 1. have elevation data, with the sorghum collection having 33 %, and rice collection, 0.44 % of accessions docuemented.

Table 1. GRIN summary of collection site information for six major crops

Crop	# collected	%with lat/long	%with locality	%with elev.
Wheat	32,612	9	48	20
Sorghum	18,551	10	72	33
Corn	16,543	17	52	28
Soybean	7,817	5	50	4
Rice	11,830	0.3	9	0.44
Forage legumes	7,459	13	60	26

```
------------------------------------------------------------------------------
--------------------
  Total:       94,812    Average:       9                      50
            20
```

Improving the information in GRIN

Information, including descriptions of site locations and habitats, does exists outside of GRIN, mainly in the *USDA Plant Inventories* , which have been published regularly since 1898. Accession information can also be found in historic index card files and acquisition logs maintained by repositories and research units, original collector notes, trip reports, correspondence and repository annual reports. Various efforts are being made to review historic documentation for inclusion into GRIN. Although substantial resources are needed to train personnel in interpreting, standardizing and entering historic information, the benefits are worthwhile. For example, efforts to upgrade the documentation of the *Lotus* (Trefoil) collection, more than doubled the amount of information in the database (Greene and West-Smith 1994). Current efforts to upgrade the *Medicago* and *Trifolium* collections tripled the number of accessions having collection site locality information (Since data has not yet been placed in GRIN, Table 1 does not reflect these efforts). As passport documentation is the responsibility of individual crop curators and repositories, varied priorities account for uneven data across the many collections within GRIN.

GIS as a management tool in genetic resource conservation

Although georeferenced data is currently limited in GRIN, the potential benefits of applying GIS tools to the conservation and management of genetic resources are providing strong incentives for curators and collection users to improve the quality and quantity of information in GRIN. Potential GIS applications can be divided into four general areas: 1) monitoring of inter- and intraspecific biodiversity; 2) developing effective acquisition strategies; 3) increasing understanding of existing collections; and 4) utilizing existing germplasm collections. The first two areas are not entirely reliant on the existence of genetic resource databases. For example: models can be developed to assess genetic erosion in primitive varieties and wild species based on the distribution of germplasm and key demographic factors across the landscape. The genetic vulnerability of cultivated crops can be assessed by examining the distribution of cultivars in a region or country. Range limitations of species can be examined in terms of known environmental or cultural factors. Potential *in situ* sites can be identified and existing genetic preserves monitored using GIS. Guarino (1995) discusses the use of GIS to develop strategies for collecting germplasm. For example, collection regions can be mapped to identify areas with desired ecogeographic attributes such as acid soils or climate extremes (Hart et al. 1996). An ecogeograhic survey using remotely sensed data can assist in acquiring a broad base sample of genetic diversity by optimizing sampling from many different environments. Potential collection areas can be identified having similar environmental envelopes as germplasm of known value.

GIS can also be used to understand and effectively use existing germplasm collections. Ecogeographic representation of collections can be assessed by overlaying collection sites on maps such as climate, soil, topography and ecosystem. Taxon range representation can be assessed by overlaying collections on species distribution maps. Accessions having minimal description of collection sites can be retroclassified using data from existing maps (Steiner and Greene 1996). Morphological descriptors or genetic variation can be linked with environmental attributes (for example: Pederson et al. 1996). Use of GIS can also assist users in selecting potentially useful material from the collection. Accessions can be identified that originate in environments having similar agroecological attributes as the target environment of the germplasm requestor. Accessions can be identified which originate from environments similar to those which have given rise to accessions with known desirable attributes such as disease and insect resistance.

Genetic resource databases as an information source

Genetic resource databases could potentially serve as an information resource to a broader scientific community, providing relatively independent global data sets to meet the unique needs of environmental modelers. For example, information from germplasm collections can be used to validate or modify agroecolgical models developed to define agricultural constrainsts and predict the agronomic potential of crop species. Based on records indicating the historic presence of the crop or wild representatives of the crop, the historic distribution and use of crop species, including adoption of new cultivars or new crop species, or continued reliance on primitive cultivars can be examined. Collection site information, including the presense of associated species, can be useful to validate predicted distributions of wild or crop species, provide micro-environment description, and document change in the landscape. Multi-disciplinary collaboration would ensure the evolution of a set of databases useful to collection curators and germplasm users, and to a broader range of professions.

The NPGS is working to enhance public-user access through the World Wide Web. This includes strengthening query capabilities to support a broader user community. Linkages between the NPGS and other national and international gene banks are being forged. For example, hyperlinks are being established among *Maize* accessions common to the NPGS and Latin American Maize Project collections. Linkages are also being established between GRIN and the Plant Genome Databases. As more institutions come online and linkages are established between common accessions, a more complete picture of germplasm collections and their contents will emerge, based upon a broad array of passport and evaluation information.

Acknowledgements

We would like to thank Dave Stout, from the USDA Western Regional Plant Introduction Station, for providing GRIN data for this paper and Mark Bohning, from the USDA National Germplasm Research Laboratory for his input and review of the manuscript.

Appendix 1. Selected list of genebank databases and related institutes

Genebanks:

1. Austria
Index Seminum Austriae, (ISA)-
Homepage contains genebank site information (addresses) and lists number of accessions by crop species. Evaluation information on 4910 accessions can be downloaded.
To access: http://www.zadi.de/igr/austria/austria.htm

2. Germany
Information Centre for Genetic Resources,(IGR), ZADI-
Homepage provides an overview of genetic resource collections in Germany, provides linkages to other national and international genebanks. Versions available in German or English
To access: http://www.zadi.de/igr/national.htm#infoeng

3. Netherlands
Centre for Genetic Resources, (CGN)-
The CGC maintains over 20,000 accessions. The extensive homepage covers genebank mission, history, institutional organization including a listing of staff and their publications, and description of various programs. General information is available on each collection, and crop-specific passport data can be downloaded. Germplasm can be requested by contacting site. Links have also been set up to the homepages of other genebanks, international agricultural organizations and related items of interest.
To access: http://www.co.dlo.nl/cgn/cgn.html

4. Scandinavia
Nordic Gene Bank, (NGB)-
Homepage contains information on collection holdings, institute mandate and organization, publication listing, including seed catalogues. There are also descriptions of programs the genebank is currently involved with.
To access:http://www.ngb.se

5. Russia
V.I. Vavilov Institute of Plant Industry, (VIR),
Homepage contains information on the institute and holdings.

6. United Kingdom
UK Plant Genetics Resources Group, (UK PGRG)-
Homepage provides access to curators of the various genetic resource centers throughout the United Kingdom. Homepages of specific institutes vary, but provide general information regarding collection contents and contact details. There are also a number of links to botanical and conservation organizations.
To access: http://nasc.nott.ac.uk:8200/home.html

7. United States:
Germplasm Resources Information Network, (GRIN),-
Central database of the USDA National Plant Germplasm System. Contains information on the NPGS (sites, staff, collection statistics) and passport and evaluation information on individual accessions, as well as seed availability and order forms.

To access:

- Internet World Wide Web: http://www.ars-grin.gov
- Internet gopher server: gopher gopher.ars-grin.gov
- Modem: dial 301-504-6227; At connect message, Type: Enter; at the login prompt, type: grin.
- PC GRIN can be downloaded from the internet or requested from the Database Management Unit

Additional information:
GRIN/Database Management Unit
USDA, ARS BA,PSI, NGRL,GRIN
10300 Baltimore Avenue
Building 003, 4th floor, BARC-WEST
Beltsville, MD 20705-2350
Phone: (301)504-5666
Fax: 301-504-5536
Email: grin@ars-grin.gov

Related institutes/URL's of interest:
1. Consultative Group on International Agricultural Research,(CGIAR)
Homepage contains recent press releases, listing of research centers and directory, newsletter and miscellaneous reports
To access: http://www.worldbank.org/html/cgiar/HomePage.html

2. International Plant Genetic Resources Institute, (IPGRI)
Homepage contains a contact list of staff and regional offices.
To access: http://www.worldbank.org/html/cgiar/directory/IPGRI.html

3. A Collection of Botany Related URL's (CBR)
Homepage contains more than 1000 botanical links, broken down into 18 subject areas which can be search using key words.
To access: http://www.helsinki.fi/kmus/botany.html

References

Cohen, J.I., Williams, J.T., Plucknett, D.L., and Shands, H.L. (1991) *Ex situ* conservation of plant genetic resources: global development and environmental concerns. *Science* 253:866-872.

Given, D.R. (1994) *Principles and practice of plant conservation* . Timber Press, Portland, OR.

Greene, S.L. and, Bohning M. (1994). Evaluation of the U.S. *Trifolium* germplasm collections: Current status. pg.74. *In* Proceedings of the 13th *Trifolium* Conference. Charlottetown, PEI, June 22-24.

Greene, S.L and S. West-Smith. (1994). Germplasm collection documentation: effective procedures and implications. p. 221. *Agronomy Abstracts*. Argonomy Society of America, Madison, WA.

Guarino, L. (1995). Geographic information systems and remotesensing for germplasm collectors. p.315-328. *In* L. Guarino, V. Ramanatha Rao, R. Reid (eds.) *Collecting plant genetic diversity: technical guidelines* Cab International, UK.

Hart, T.S., S.L. Greene, A. Afonin. (1996) Mapping for Germplasm Collections: Site Selection and Attribution. *Proceedings of the third international conference on integrating GIS and environmental modeling*. NCGIA, Santa Barbara, CA, WWW and CD.

National Research Council. (1993) *Managing global genetic resources: agricultural crop issues and policies* . National Academy Press, Washington D.C.

Pederson, G.A., T.E. Fairbrother, and S.L. Greene (1996). Cyanogenesis and climate relationships in U.S. white clover germplasm collection and core subset.*Crop Sci.*(in press).

Steiner, J.J. and S.L. Greene. (1996) Proposed ecological descriptors and their utility for plant germplasm collections. *Crop Sci.*36: (in press).

Steiner, J.J. and C.J. Poklemba. (1994) *Lotus corniculatus* classification by seed globulin polypeptides and relationship to accession pedigrees and geographic origin. *Crop Sci.* 34:255-264.

White, G.A., H.L. Shands, G.R. Lovell. (1989) History and operations of the Natioanl Plant Germplasm System. p. 5-56. *In* J. Janick (ed.) *Plant Breeding Reviews vol. 7. The National Plant Germplasm System of the United States* . Timber Press, Inc. Portland, OR.

Williams, J.T. (1989) Plant germplasm preservation: A global perspective. p. 81-96. *In* L. Knutson and A.K. Stoner (ed.) *Biotic Diversity and Germplasm Preservation, Global Imperatives* . Dordrecht, Netherlands: Kluwer Academic Publishers.

Author Information

Stephanie L. Greene
Forage Legume Curator, USDA, ARS National Plant Germplasm System
Washington State University-Prosser
24106 N. Bunn Road, Prosser, WA 99350
Phone: 509-786-9265
Fax: 509-786-9370
e-mail: sgreene@ars-grin.gov

Thomas Hart
GIS Specialist, Spatial Data Associates
20 Bush Lane, Ithaca, NY 14850
Phone: 607-257-0951

Fax: 607-257-3167
e-mail: jhart@lightlink.com

Kris C. Matson, John E. Fels

Approaches to Automated Water Table Mapping

Water table depth, the distance from land surface to the top of the surficial aquifer, is an important map layer, parameter, and input variable for a wide variety of environmental models and decision support methodologies: maps of the water table are used to estimate ground water flow direction and velocity; regional water table maps are important components of ground water vulnerability assessments; ground water depth is an important variable in many hydrologic and pollution transport models; and regional maps of ground water depth are commonly used in environmental decision making such as locating landfills and wastewater disposal sites.

Because water table depth is measured at point locations (in wells or soil borings), water table mapping requires hydrogeologically appropriate techniques to generalize the point measurements. The creation of water table maps is a well accepted practice in ground water investigations - water levels are measured in a network of wells and a surface is interpolated between measuring points. However, this practice works only if the investigated area is small or numerous measuring points are available. Methods are not well established for mapping water table depth over large areas where measurements are sparse and opportunistic.

As part of a ground water vulnerability mapping project for the State of North Carolina, three approaches were considered for regional water table mapping. The first approach uses deterministic modeling, the second uses statistical modeling, and the third uses landscape classification. The three approaches are described in this paper; implementation of the landscape classification method chosen for mapping is discussed in the conference paper, "A Cognitively-based Approach for Hydrogeomorphic Land Classification using Digital Terrain Models" (Fels and Matson 1996).

1.0 Introduction and Context

1.1 Background

Ground water, by definition, is hidden from view beneath the land surface. Monitoring wells provide the only direct means of observing ground water. Virtually all we know about ground water is a product of applying and testing hypotheses to generalize observations and measurements conducted in the field or in the laboratory. Rules, models, or schemas for spatially and temporally generalizing monitoring (sample) data across the ground water system are inherent and essential to hydrogeologic science. Our understanding of ground water is the product of a long history of hypothesis and model development, testing, and refinement.

Models employed in hydrogeology span a range of sophistication, abstraction, and rigor, from conceptual models of general hydrogeologic relationships, to deterministic models developed from partial differential equations representing conservation of mass and energy in the ground water system. Any spatial representation of ground water properties or phenomena requires models of some sort to spatially and temporally generalize points measurements.

The water table is the surface of saturated conditions, below which all pores and voids in soil, sediment, or rock are completely filled with water (Heath 1989). Depth to the water table is measured in wells screened across the zone of water table fluctuation (Figure 1). The position of the water table is the product of a wide range of static and dynamic environmental conditions and processes affecting the rate at which water enters and leaves the saturated zone. The water table rises if the rate of water added (recharge) exceeds the rate of water leaving (discharge); conversely, the water table falls if discharge exceed recharge. The water table surface is therefore not static, nor flat (as the name implies), but responsive to climatic, vegetative, geomorphic, and geologic conditions.

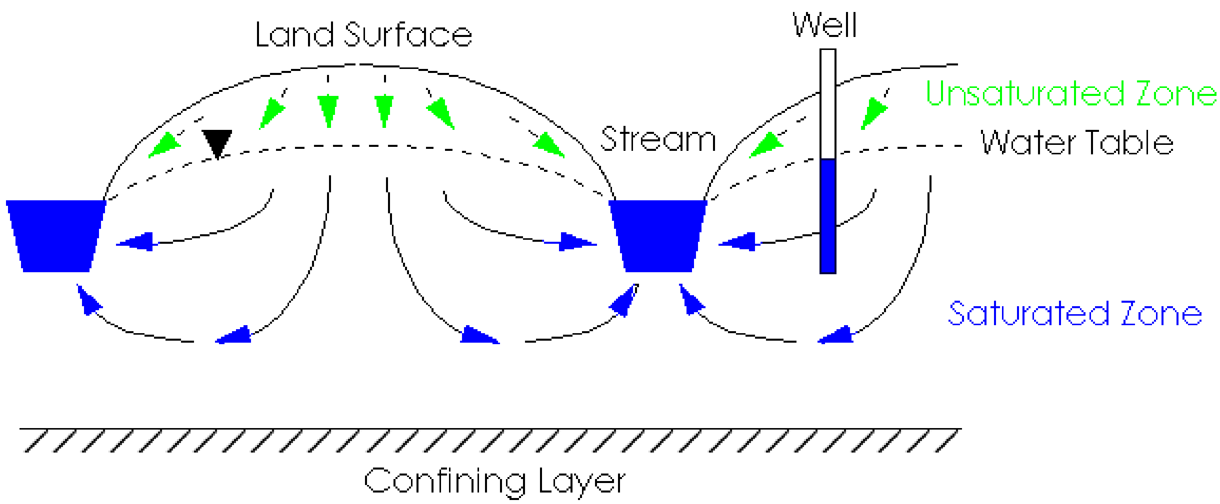


Figure 1. Schematic cross-section of a typical surficial aquifer system in humid regions. Solid arrows represent saturated flow, dashed arrows represent unsaturated flow (After Freeze and Cherry 1979).

Traditional water table mapping uses graphical methods to interpolate between water table measurements and hydrogeologic boundaries, with professional judgment and experience filling the gaps in sampling. Computer assisted approaches may incorporate surface mapping methods such as trend surface interpolation and kriging; many of these tools are currently provided in GIS software. Other methods employ mathematical modeling to predict water table elevation from hydrogeologic conditions and processes.

Surface interpolation and mathematical modeling require a high degree of site-specific knowledge and observations. Although detailed water table information may be available in site-specific investigations, sparse and opportunistic water table information must frequently be used in regional hydrogeologic assessments. Interpolation methods do not work well in these situations because they define only general (large wavelength) changes in water table elevation, missing the local (short wavelength) variations in water table elevation with respect to land surface. New developments in spatial analysis and the availability of digital environmental data for large areas provide a number of possible solutions to overcome these regional mapping problems.

As part of a statewide assessment of ground water vulnerability using the DRASTIC method (Aller *et al.* 1987), the Groundwater Section in the North Carolina Department of Environment, Health, and Natural Resources, required a state-wide map of average annual depth to the water table. The goal of this effort was to use available information to develop a predictive map of average annual water table depth, at a scale of 1:250,000, in a short time-frame, and with a limited budget. Project staff considered using a GIS combined with either deterministic models, statistical models, or landscape classification models in order to accomplish this task. The issues involved in implementing these various models in a GIS are discussed below. The implementation of the landscape classification approach chosen for state-wide water table mapping is discussed separately in the conference paper "A Cognitively-based Approach for Hydrogeomorphic Land Classification using Digital Terrain Models" (Fels and Matson 1996). References in this paper to GIS and spatial analysis assume raster, or grid-cell systems.

1.2 Conceptual Model of Water Table Depth in the Study Area

Nearly 100 years ago King (1899, as cited in Domenico and Schwartz 1993) observed that the water table is a subdued replica of the land surface in humid regions (Figure 1). This conceptual model is among the oldest in hydrogeology. In North Carolina, water table depth varies widely from just beneath land surface in much of the Coastal Plain to over 100 feet beneath high peaks in the Blue Ridge Mountains. The water table is expressed at land surface where it discharges into rivers, streams, canals, lakes, wetlands, and other surface water bodies throughout the state. At a coarse level, Heath (1988) divides ground water occurrence in North Carolina into two regions of distinct hydrogeologic conditions: the Coastal Plain, and the Piedmont and Blue Ridge Mountains (Figure 2).

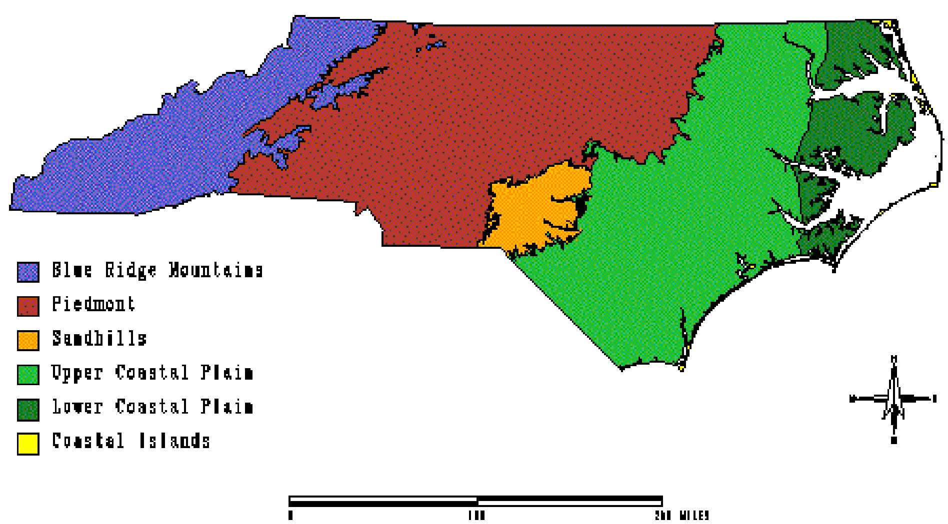


Figure 2. The major physiographic regions of North Carolina. Each region has different hydrogeologic characteristics.

The Coastal Plain hydrogeologic region is characterized by sequences of sedimentary deposits overlying metamorphic and igneous bedrock (Heath 1988). The landscape is relatively flat and low in elevation, shallowly dissected by a network of rivers and streams. The Coastal Plain may be further divided on the basis of geomorphology and geologic materials into the Sandhills, the Upper Coastal Plain, the Lower Coastal Plain, and the Coastal Islands (Figure 2). In the Lower Coastal

Plain and Coastal Islands, the water table is at or very close to the land surface; wetlands and hydric soils are particularly abundant. The water table is deeper (approximately 5 to 15 feet) beneath modern and relict sand dunes, and on sandy scarps and ridges. Upland areas in the Sandhills and the western Upper Coastal Plain have the deepest water levels although it is unusual for depths to exceed 20 feet. Water levels are shallowest close to streams, in floodplains, and on wide, flat interstream divides (Daniels *et al.* 1984).

The Piedmont and Blue Ridge Mountain regions are hilly to mountainous and underlain by saprolite and alluvial sediments over bedrock (Heath 1988). Ground water occurs in both the sediments and underlying fractured bedrock. Wetlands are less common than in the Coastal Plain but can occur near streams and in floodplains where the water table is close to or at the land surface. The depth to the water table deepens at the edge of floodplains, where the slope of the land surface increases and alluvial materials give way to colluvial sediments and saprolite. Saprolite, sediment derived from the in-situ weathering of underlying bedrock, is generally thickest at the tops of hills and ridges in the Piedmont and the water table is deepest in these locations at approximately 40 feet (Daniel 1989). In the Blue Ridge Mountains, alluvial sediments are generally coarser, slopes are generally steeper, and saprolite is thin or absent on the slopes, ridges, and mountains. Water table depths are shallow near streams but deepen with increasing slope and topographic relief. Water table depths on inhabited hills and ridges where wells are installed average 60 feet (Daniel 1989) but depths beneath high narrow ridges and peaks exceed 100 feet.

1.3 Data Sources

Water table mapping requires a combination of hydrogeologic and ancillary geographic information. Quality controlled, hydrogeologic information available to assist in North Carolina mapping include ambient ground water monitoring data, drilling logs, stream gauge data, and hydrogeologic reports. Digital geographic data for North Carolina include: 1:250,000 scale DEMs, 1:100,000 scale streams and surface water bodies, 1:250,000 scale ground water discharge areas, 1:1,000,000 scale ground water recharge rates, 1:250,000 scale Physiographic Provinces, 1:250,000 scale geology, and a 1:500,000 scale hydrogeologic map of the Piedmont and Blue Ridge Mountains.

2.0 Deterministic Modeling Approach

Deterministic approaches to ground water modeling employ mathematical equations to explicitly represent the physical relationships, behavior, and properties of the ground water system. Deterministic models use the ground water flow equation, a second-order partial differential equation incorporating Darcy's law of fluid flow and the equation for conservation of fluid mass for three-dimensional flow through porous media (Freeze and Cherry 1979). Governing equations are derived from this equation by making assumptions about the hydrogeologic properties of the system under investigation. By defining boundary conditions, these equations are simplified and solved analytically or numerically. Model results are sensitive to the assumptions, parameters, and boundaries by which the system is defined. The accuracy of mathematical modeling methods are therefore inherently constrained by knowledge of the hydrogeologic parameters and the reasonableness of the simplifying assumptions.

Deterministic water table mapping in a GIS requires linking an analytical or numerical model to calculate a water table elevation surface. This surface must be subtracted from land surface elevation to create a map of depth to the water table.

2.1 Analytical models

Analytical ground water models employ closed-form mathematical solutions to the ground water flow equation (Freeze and Cherry 1979). One such solution is the water profile equation (Equation 1, Jacob 1943) which may be used to determine hydraulic head, *h* in an aquifer between two surface water bodies:

$$h = \frac{w}{T}\left(ax - \frac{x^2}{2}\right)$$

(1)

Where *h* is the height of the water table above the bounding surface water bodies, *w* is recharge to the aquifer, *T* is the aquifer transmissivity (hydraulic conductivity times aquifer thickness), *a* is the distance from a surface water body to the ridge line (or one-half the total distance between water bodies), and *x* is the horizontal distance to the nearest surface water body.

Key assumptions made to arrive at this solution are that: groundwater flows only in a horizontal direction perpendicular to the surface water bodies, the aquifer is infinite in extent parallel to the surface water bodies and homogeneous and isotropic in its material properties, and the surface water bodies are parallel and at constant and equal elevation (the Dupuit-Forchheimer assumptions for unconfined, saturated ground water flow (Freeze and Cherry 1979)). These boundary conditions and assumptions are most closely approached on North Carolina's coastal barrier islands. C. E. Jacob (1943) first developed and employed this model to simulate the water table profile on Long Island, New York.

Implementation of this model in GIS for average annual water table mapping requires surface water hydrography to determine boundary conditions and digital elevation data to determine surface water elevation, the datum from which water table elevation (*h*) is calculated. To remain consistent with model assumptions, average annual recharge and aquifer conductivity are constant throughout the mapped region. The height of the water table is determined for a particular grid cell by calculating its distance from the nearest surface water body, the total distance between surface water bodies at that point, and then applying the water profile equation through grid algebra or database calculations. The model is calibrated by comparing field observations with model results and adjusting hydraulic conductivity or recharge rates within acceptable limits. In the absence of either recharge or conductivity data, inverse modeling may be employed to determine these parameters from field measurements of the water table.

2.2 Numerical models

Numerical models employ algebraic approximations to solve ground water flow equations. Finite difference numerical methods discretize the continuous ground water flow field into discrete blocks, or elements, within which hydrogeologic properties are assumed constant. The cell-based discretization used in finite-difference methods allows coupling of numerical models with raster GIS software.

Ground water flow in a homogeneous, isotropic aquifer with hydraulic conductivity *K*, under conditions of uniform recharge *w*, and assuming horizontal flow is represented by Equation 2 (Domenico and Schwartz 1990):

$$\frac{\partial^2 h^2}{\partial x^2} + \frac{\partial^2 h^2}{\partial y^2} = -\frac{2w}{K}$$

(2)

where *h* is water table height and *x* and *y* are the horizontal dimensions. Letting *u = h^2*, the numerical approximation for Equation 2 is (Domenico and Schwartz 1990):

$$u_{i,j} = \frac{1}{4}\left(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}\right) + \frac{2w\Delta x^2}{K}$$

(3)

where *i,j* denotes a particular cell in row *i* and column *j* in a regular grid with cells of equal *x* and *y* dimensions, and *delta x* is the grid cell width. Boundary conditions must be specified as *h^2*. Equation 3 calculates hydraulic head as the average of the four adjacent cells plus the constant term relating recharge and conductivity.

Equation 3 can be implemented in a raster GIS as a customized digital filter. Boundary conditions, which must remain fixed (i.e., unfiltered), are set as the squared elevations of surface water bodies. The remainder of the grid is given an intermediate value and Equation 3 is iteratively applied until succeeding values of calculated head converge within a predetermined limit (Domenico and Schwartz 1990). Model results are calibrated from field observations of water level and the conductivity and recharge parameters are adjusted within acceptable ranges.

2.3 Discussion -- Deterministic Models

Deterministic models provide the advantage of explicit solutions to water table elevation, but they are not straightforward to implement in a GIS. Deterministic model applications require adherence to the assumptions made to arrive at solutions to the ground water flow equation. Although the specification of boundary conditions (surface hydrography) is relatively straightforward, model assumptions would most likely be violated for regional mapping because of the spatial variability of hydrogeologic parameters in the surficial aquifer. Deterministic models are usually implemented on site-specific scales where a high level of control is possible for assumptions, parameterization, and calibration. Numerical models are probably best implemented using customized digital filtering techniques in image processing software. The requirement for tuning the model through calibration is essential, but difficult to implement in GIS over large areas with many data points. Incorporating the spatial variation in aquifer properties through tightly coupled GIS and deterministic model linkages requires such substantial overhead in algorithm development that most researchers find loosely coupled linkages to external deterministic models more appropriate (Stuart and Stocks 1994). Nonetheless, considerable potential exists for developing better methods of coupling deterministic models with GIS (Benjamin Houston 1995, personal communication).

David Maidment (1996, personal communication) reports success in implementing equation 3 in a GIS as a linked ground water/surface water model. For the state-wide water table mapping exercise at hand, the development time required to establish model linkages and prepare the data sets was prohibitive.

3.0 Statistical Modeling Approach

Although the water table is a subdued replica of land surface, the relationship between the two surfaces is not simple, but subject to terrain, geology, climatic conditions, soil properties, and human influences. These relationships are difficult to incorporate and quantify deterministically on a regional scale. A statistical modeling approach to water table mapping seeks to determine which environmental variables are the best predictors of water table depth without explicitly characterizing the physical processes involved, and establishes a mathematical function for the relationships. In statistical models, the explanatory variables may be causative or simply useful estimators of the response variable (Clarke 1994); readily measured environmental characteristics may thus serve as surrogates for hydrogeologic variables in a statistical model. Similar techniques have been employed by Fels (1994) to predict vegetative diversity, and by Gessler *et al.* (1995) to predict soil properties, from terrain variables extracted from DEMs.

3.1 Statistical Models

Developing statistical models for water table mapping begins with conceptualizing relationships between water table depth, the response variable, and other mappable characteristics, the explanatory variables. In the North Carolina Piedmont and Mountains, water table depth generally increases with increasing distance from streams as well as with increasing relative and absolute elevation Daniel (1989). In the Coastal Plain, water table depth increases with increasing relative elevation and distance from streams and then it shallows to the land surface on wide interstream divides (Daniels *et al.* 1984). One or more terrain measurements, or geomorphometrics, may thus serve as explanatory variables for water table depth.

Following the notation of Clarke (1994) the general form of a statistical model for water table depth may be written as:

$$WT_s = f\left(\{Elev\}_s,\ \{Slope\}_s,\ \{Dist\}_s,\ \ldots\ ;\ \theta\right) + \epsilon_s$$

(4)

where *WT* is water table depth; *s* is the spatial domain of the function; *Elev, Slope, Dist, etc.* are the explanatory measurements or values we can extract from continuous data sets in the model domain, such as elevation, slope, and distance to stream; *theta* is the set of fixed parameters which are estimated from the data; and *epsilon* is the random component representing fluctuations about the systematic component, *f (...)*. The goal of model development is to determine the most appropriate form of the function *f (...)*, the best estimates for the parameters *theta*, and the estimated probability distribution of *epsilon* (Clarke 1994).

The simplest form of the statistical model is the case in which Equation 4 is univariate and linear in the parameters. For example, the simplest form of a statistical model for water table depth could be one in which depth is explained by elevation (*Elev*) within an acceptable distribution of *epsilon*:

$$WT = \alpha + \beta(Elev) + \epsilon$$

(5)

Equation 5 is a linear regression equation in which the parameters *alpha* and *beta* are constants. Inspection of Equation 5 reveals that water table depth (*WT*) will be zero for only one value of elevation (negative *alpha/beta*). This simple hypothetical model can only hold in a restricted spatial domain where surface water bodies are all at the same elevation. Wider model application and performance would likely be achieved with the incorporation of additional explanatory variables and the use of more complex model forms such as quadratic, logarithmic, and gaussian.

To develop a statistical model for water table depth and implement it in a GIS, the largest possible set of explanatory spatial variables should be assembled for locations with water table measurements. The explanatory variables initially considered may include any number of geomorphometrics, including elevation, slope, distance to stream, and meso- and micro-scale landform. Data visualization is useful in determining possible explanatory variables. The set of water table measuring locations must be fully representative of the spatial domain: an assumption is made as to which variable explains most of the water table variation and sample locations are randomly distributed within strata of this variable. Statistical methods are employed to determine the form of the equation and estimate the parameters for each of the explanatory variables and the random component. This results in a model equation which directly predicts water table depth.

The model is implemented in a GIS by calculating the relevant explanatory variables for all cells in the model domain. The model equation is applied using grid algebra or database functions to calculate water table depth for each grid cell from the explanatory variables in the same location.

3.2 Discussion -- Statistical Models

The statistical approach to water table mapping can take into account more of the spatial variability in average water table depth on a regional scale than is generally possible with deterministic models. Statistical models are also rather simple to couple and implement in a raster GIS. On the other hand, existing water level data points may be inadequate or insufficiently stratified for model development, requiring additional monitoring locations to be established, an important restriction if only opportunistic data is available. Other than experimental applications in Mississippi by the US Geological Survey (Chuck O'Hara 1994, personal communication), the statistical approach to water table mapping has not been widely applied. The greatest limitation of statistical models, and the strength of deterministic models, is in simulating the response of the water table to short-term environmental stresses. Statistical modeling is useful for mapping long-term averages but it is generally not adequate for modeling events such as drought, excessive precipitation, or pumping in a well field (David Evans 1996, personal communication).

4.0 Landscape Classification Approach

The previous methods generate a continuous surface of average annual water table elevation or depth. This is the most widely useful spatial representation of the water table because it allows determination of the ground water gradient, an important parameter or variable in many models and hydrologic assessments. However, another mapping approach is the classification of the landscape into ranges of average annual water table depth.

Although land and water table surfaces are highly complex continua, they may be classified into regions of less variability. In the same manner that pedologists use directly observable geomorphic, climatic, geologic, and vegetative characteristics to interpret and map soil properties (Daniels and Hammer 1992), hydrogeologists use similar environmental clues to predict water table depth in the field. This approach seeks to map the water table indirectly by classifying discrete landforms with predictable ranges of average annual water table depth. The resulting classification model is an algorithmic formalization of the conceptual models used in traditional field mapping.

4.1 Landscape Classification Models

Classification requires rules for spatial discrimination; automated classification requires codable rules for discriminating cells in gridded data sets. A variety of measurements may be calculated from a DEM to represent the landscape characteristics of a particular cell. These measurements, called geomorphometrics, are local, focal, and zonal functions which describe the geometry and topology of cells in the DEM. Landscape settings, or land types, may be defined by logically combining these metrics. In some cases, land types may be defined with a single metric -- cliffs are regions in which slopes exceed a particular value. In other cases, a combination of metrics is required -- ridges have low slope values, but they occur at high relative elevations, and although they are continuous features, elevation may change considerably along their length. The boundaries of land types are often subjective. For example, at what slope value does a ridge become a side-slope? R. Dikau (1989) seeks to define a comprehensive geomorphic taxonomy based on geomorphometrics for landscape features of various scales and for various purposes. In this approach, we seek to define only those mesoscale landforms (i.e., ridges, slopes, valleys, etc.) which reflect water table depth classes.

The classification model is developed in three steps: first, the thematic classes of water table depth are specified; then the land types which reflect these classes are determined; lastly, the land types are interactively classified from a DEM.

For the current study, a hierarchical, 2-level, land type classification scheme was developed to reflect the depth to water table ranges in the DRASTIC methodology (Aller *et al.* 1987). Within each of North Carolina's six physiographic regions (Figure 2), land types were selected to reflect the prescribed depth ranges (Fels and Matson 1996). Water table depth ranges chosen for the thematic classes should be selected on the basis of the planning or modeling need at hand (the DRASTIC ranges do not necessarily reflect a relevant discrimination of water table depth). It was found that logical combinations of two metrics, slope and landscape position index (a focal function which measures the relative position of a cell with respect to its surrounding cells (Fels 1994)), allow discrimination of the mesoscale land types in the classification scheme. For example, in the Sandhills all cells with landscape position values less than -.005 were classified as hills and ridges (landscape position index is positive in valleys, neutral in mid-slope position, and negative on ridges); of the remaining cells, those with slope values of 2% or greater were classified as sideslopes and those having slopes less than 2% were classified as flats (Fels and Matson 1996). The more diverse features in the Piedmont and Mountains required a more complex classification procedure using the same metrics. Deciding where to break land type classes requires repeated parameterization and visualization of the results. Landscape position and slope values are varied until the best (most cognitively acceptable) fit is achieved between the classified land types and knowledge of water table depth.

4.2 Discussion -- Landscape Classification

Classified depth to water table ranges offer a valuable means of discriminating land areas for certain planning, assessment, and modeling purposes. Although a thematic map layer is not as widely useful as a continuous surface, landscape classification allows water table mapping without the resources required by the deterministic and statistical approaches. Landscape classification can assist field scientists by extending their knowledge into new areas using visual representations of conceptual models for ground water occurrence. Because this approach does not make explicit use of water table measurements, classification accuracy is subject to the knowledge of water table and geomorphic relationships by those conducting the mapping. Classification accuracy is readily measured by cross-tabulating water table depth measurements with the classified ranges and statistically analyzing the amount of agreement. The results of this analysis may be used to adjust the algorithm by which one or more landtypes are discriminated.

5.0 Conclusions

Automated regional water table mapping is a complex exercise. The three approaches discussed here illustrate both the value and limitations of using GIS for the task. The most significant value of GIS is the ability to use extensive environmental data sets to parameterize models and establish boundary conditions over large areas. Although some of the tools required to conduct water table mapping are available in raster GIS or image processing packages, all three approaches require customized software for data analysis, model linkages, and export to statistical packages.

Deterministic models offer explicit definition of a water table surface from physical principles and the ability to simulate the response of the water table to short-term environmental stresses. Deterministic models are difficult, however, to implement reliably at regional scales. Statistical models permit direct prediction of water table depth from readily measured explanatory variables without the need to characterize the physical properties or processes in the ground water system. Statistical models also incorporate explicit measures of accuracy, an important (but often missing) component in mapping and modeling activities. Both deterministic and statistical models are data intensive, and generally require new data collection for model development, implementation, and calibration. Landscape classification is an intuitively appealing and efficient approach which formally characterizes conceptual models for water table depth in terms of geomorphic features. Although classification accuracy is measurable, it is subject to the hydrogeologic and geomorphic knowledge of those developing the model algorithms.

Landscape classification was chosen for initial state-wide water table mapping in North Carolina because of its short development time, low implementation cost, and the ability to incorporate existing hydrogeologic knowledge. Classification of land types by hydrogeologic significance supports future model applications by providing a structure for stratifying monitoring locations.

Regional water table mapping is a promising area of research in the geographic and hydrogeologic sciences. The mapping methods presented here will have greater predictive accuracy as they are implemented, compared, and refined in other locations. Further work will also lead to new methods of quantifying regional ground water properties, enhancing the protection of ground water resources. We hope this discussion encourages others to apply and evaluate these water table mapping approaches in their own work.

6.0 Acknowledgments

The authors wish to thank the Groundwater Section of the North Carolina Department of Environment, Health, and Natural Resources, and the United States Environmental Protection Agency, for the opportunity and funding to conduct this work. We thank Carl Bailey, Ted Mew, Arthur Mouberry, and Perry Nelson for conceptualizing and supporting efforts to characterize the hydrogeology of the shallow aquifer system in North Carolina. We also thank Ralph Heath, consultant, and Bruce Lloyd, Charles Daniel, Tim Spruill, and Gerry McMahon of the United States Geological Survey for their insights into regional water table mapping and the approaches discussed here. We are also indebted to Scott Huffman and Kai-Ping Wang of the Groundwater Section for their technical support in this mapping project. Lastly, we are grateful to David Evans, Casson Stallings, and Benjamin Houston for their review of an earlier manuscript of this paper.

7.0 References

Aller, L., Bennet, T., Lehr, J.H., Petty, R.J., and Hackett, G. (1987) *DRASTIC: A Standardized System for Evaluating Ground Water Pollution Potential using Hydrogeologic Settings*, US EPA Document #EPA/600/2-85-018. Washington, DC: US Government Printing Office.

Clarke, R.T. (1994) *Statistical Modelling in Hydrology*. New York: John Wiley & Sons.

Daniel, C.C. (1989) *Statistical Analysis Relating Well Yield to Construction Practices and Siting of Wells in the Piedmont and Blue Ridge Provinces of North Carolina* (US Geological Survey water-supply paper 2341-A). Washington, DC: US Government Printing Office.

Daniels, R.B., Kleiss, H.J., Buol, S.W., Byrd, H.J., and Phillips, J.A. (1994) *Soil Systems in North Carolina* (NC ARS Bull. 467). Raleigh, NC: North Carolina State University.

Daniels, R.B., and Hammer, R.D. (1992) *Soil Geomorphology*. New York, NY: John Wiley & Sons.

Domenico, P.A. and Schwartz, F.W. (1990) *Physical and Chemical Hydrogeology*. New York, NY: John Wiley & Sons.

Dikau, R. (1989) The Application of a Digital Relief Model to Landform Analysis in Geomorphology, in Raper, J. (ed.) *Three Dimensional Applications in Geographic Information Systems*. London: Taylor & Francis.

Evans, D. (1996) North Carolina State University, Dept. of Marine, Earth, and Atmospheric Sciences, personal communication.

Fels, J.E. (1994) Modeling and Mapping Potential Vegetation Using Digital Terrain Data: Applications in the Ellicot Rock Wilderness of North Carolina, South Carolina, and Georgia (Ph.D. Dissertation). Raleigh, NC: North Carolina State University.

Fels, J.E. and Matson, K.C. (1996) A Cognitively-based Approach for Hydrogeomorphic Land Classification using Digital Terrain Models. *Proceedings, Third International Conference on Integrating GIS and Environmental Modeling*. Santa Barbara, CA: National Center for Geographic Information and Analysis. WWW, CD.

Freeze, R.A., and Cherry, J.A. (1979) *Groundwater*. Englewood Cliffs, NJ: Prentice-Hall.

Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J. (1995) Soil-Landscape Modelling and

Spatial Prediction of Soil Attributes. *Int. J. Geographical Information Systems* 4: 421-432

Heath, R.C. (1989) *Basic Ground-Water Hydrology* (US Geological Survey water-supply paper 2220). Washington, DC: US Government Printing Office.

Heath, R.C. (1988) Hydrogeologic Settings of Regions, in Back, W., Rosenshein, J.S., and Seaber, P.R. (eds.) *Hydrogeology*. Boulder, CO: Geological Society of America.

Houston, B. (1995) North Carolina State University, Dept. of Marine, Earth, and Atmospheric Sciences, personal communication.

Jacob, C.E. (1943) Correlation of Ground-Water Levels and Precipitation on Long-Island, New York. *Transactions, American Geophysical Union, Papers--Hydrology*, pp. 564-573.

King, F.H. (1899) Principles and Conditions of the Movements of Groundwater. *US Geological Survey 19th Annual Report, Part 2*, pp. 59-294.

Maidment, D.R. (1996) University of Texas at Austin, Dept. of Civil Engineering, personal communication.

O'Hara, C. (1994) US Geological Survey Water Resource Division, Jackson Mississippi, personal communication.

Stuart, N., and Stocks, C. (1993) Hydrological Modelling Within GIS: An Integrated Approach, in Kovar, K., and Nachtnebel, H.P. (eds.) *HydroGIS 93: Application of Geographic Information Systems in Hydrology and Water Resources* (Proceedings of the Vienna Conference, April, 1993). Wallingford, UK: IAHS Press.

Kris C. Matson
Research Associate
North Carolina State University
Dept. of Biological and Agricultural Engineering
Box 7625, Raleigh, NC 27695-7625
Ph: 919-515-6792, Fax: 919-515-6772
matson@eos.ncsu.edu

John E. Fels
Research Associate Professor
North Carolina State University
Dept. of Landscape Architecture
Box 7701, Raleigh, NC 27695-7701
Ph: 919-515-7341, Fax: 919-515-7330
fels@unity.ncsu.edu

Deliniation of Chemical Hydrological Response Units (CHRUs) within a GIS for hydrochemical modeling in the mesoscale Broel catchment in Germany

Ulrike Bende-Michl
University of Jena
Department of Geography
Loebdergraben 32
07743 Jena
Germany

Abstract

Characterizing and modeling hydrochemical catchment dynamics is dealing with the challenge of upscaling point measurements to a regional, spatially distribution. Additionally regional spatial dynamics of nonpoint and point sources and their relation to catchment management have to be considdered regarding water qualtiy concerns. Theese two challenges can be met by applying the concept of Chemical Hydrological Response Units (CHRUs) and using GIS-analysis to delinate their spatial distribution. CHRUs can be defined as computational elements having similar hydrochemical dynamics depending on their landuse management and physiographic catchment characteristics. The concept was applied to the mesoscale Broel catchment in Germany (A=216 skm). The landuse pattern of the catchment was mapped in the field, digitized by the aid of a GIS-system and compared to remote sensed data. A project database included analysed water samples and atmospheric deposition samples for dissolv! ed solids, fertilzer application a nd landscape catchment characteristics. Solute chemical balancing within the watershed in relation to the CHRU-concept were than compared to the measured catchment output at the basin outlet. Results showed deviations associated with undetected point sources but also fits during periods with unique hydrological conditions. In summary the concept of CHRUs by using GIS is a powerful and practical method to evaluate the hydrochemical dynamics and the estimation of their environmental hydrochemical changes in a heterogenuous catchment. Therefor it can be used for longterm observation of water quality within rivers. Future research will combine the CHRU-method with a hydrological model like MMS/PRMS and coupling it with the AgNPS and CHRIS model.

Baxter E. Vieux[1], Ph.D., P.E., and Nadim S. Farajalla[2], Ph.D.

Temporal and Spatial Aggregation of NEXRAD Rainfall Estimates on Distributed Storm Runoff Simulation

ABSTRACT

NEXRAD is a newly-deployed observational platform for high resolution measurement of rainfall and other products using Doppler radar by the National Weather Service. New observations in the hydrologic sciences offers insight into processes including spatial and temporal variability of one of the most important inputs to a hydrologic model: rainfall. As enhanced estimates of rainfall in space and time become available, questions are raised pertaining to spatial and temporal variability and the influence on distributed hydrologic modeling. A GIS is used to help answer which time and space intervals capture the variability essential for accurate and reproducible simulations of a particular river basin.

The effect of spatial aggregation of the rainfall maps on runoff simulation is investigated by aggregating the NEXRAD radar data from a 0.5 to 6 km and simulating the impact of spatial resolution on storm runoff simulation in a 1200 km² river basin located in Oklahoma. GIS is used to process rainfall estimates derived from radar reflectivity. Besides the role in processing the data, GIS and an internally integrated hydrologic simulation model ***r.water.fea*** are used to simulate storm runoff using distributed maps of rainfall and other parameters controlling the hydrologic process of storm runoff.

INTRODUCTION

Radar rainfall estimates are useful for flood forecasting because the radar image provides spatial patterns of rainfall, whereas rain gages are only point measurements. A recently deployed Doppler radar system covering much of the U.S. is the result of a joint program undertaken by the U.S. Departments of Commerce (DOC), Defense (DOD), and Transportation (DOT). This program is commonly known as NEXRAD (Next Generation Weather Radar) or Weather Surveillance Radar-1988 (WSR-88D). Besides severe storm and meteorological forecasting, the WSR-88D radar is an important source of rainfall information to hydrologists for real-time flood forecasting as well. The objective of this paper is to present the use of GIS in the processing of WSR-88D radar reflectivity data for flood forecasting of the 1200 km² Blue basin in south central Oklahoma.

Hydrologic simulations require rainfall information as point or areal estimates in space and at various time intervals depending on the type of analysis. Water resources forecasting may rely on monthly or annual precipitation whereas flood forecasting and hydrologic design may require rainfall rates at minute, hourly or by event time scales. River basins often require some estimate of the spatial distribution of rainfall over spatial scales of several to thousands of square kilometers. Rain gages are often too far apart in space causing significant errors in the estimation of the spatial extent of the rainfall. Insufficient temporal intervals, such as daily accumulation when hourly is required by the model, are additional problems.

Resolution in time and space of rainfall rates in relation to the hydrologic processes operating at the scale of interest, from catchment to river basin, affects the accuracy of the rainfall estimates and therefore, the hydrologic simulation. Radar is an important source of rainfall rates distributed in space and time not provided by rain gages. Both lumped and distributed hydrologic modeling require accurate estimates of the spatial and temporal variability of the rainfall as input. Radar estimates of rainfall rate in space and time are essential to distributed hydrologic modeling, such as river stage, flood forecasting, hydropower, and reservoir gate operation.

GIS plays an important role in the management and processing of spatial data, and in this research, radar estimates of rainfall. Further, since the flood forecasts are made using an internally integrated GIS hydrologic model, ***r.water.fea*** (Vieux, Farajalla, and Gaur ,1993; and Vieux and Gaur, 1994), GIS is used both for data ingest and for hydrologic modeling. The following sections describe the ingest, processing, scale and resolution of WSR-88D radar data in hydrologic simulations using the raster GIS, GRASS.

WSR-88D Precipitation Products

A brief description of the radar data acquisition and processing stream is necessary to understand the source, quality, and type of data available for ingest to a hydrologic model. The WSR-88D Doppler radar is an S-band transmitter that records the reflectivity, radial velocity and spectrum width of reflected signals. This is the WSR-88D base data produced by the Radar Data Acquisition (RDA) unit. Data at this stage is referred to archive level II. The next stage in the processing stream is the Radar Product Generator (RPG) which applies computer algorithms to produce meteorologic and hydrological products. Data from this stage is referred to archive level III. A more complete description of these data products and processing may be found in Crum and Alberty (1993). The unprocessed reflectivity obtained from level II data is converted to rainfall rates using a relationship between reflectivity and rainfall rate called the Z-R relationship. The base reflectivity is collected in a radial coordinate system during 360 degree azimuthal scans in ranges of 1 km and sectors of one degree. Each scan is incremented 0.5 degree in elevation to form a volume scan. Depending on the number of tilt angles used, a volume scan takes from 5 to 10 minutes to complete. During the RPG processing, the data is aggregated in time and space depending on the product generated. The precipitation products generally available to the public are at an aggregated resolutions and time intervals. The NWS River Forecast Centers (RFC) forecast river stage using a mosaic of radars aggregated to 4 km and 6 hour spatial/temporal resolution. This product is known as stage III which is used by the RFC to forecast river stage.

The research presented herein describes a Very Fine Grid (VFG) precipitation processing algorithm which does not aggregate, thus preserving space and time variability. While aggregation may reduce noise and other errors associated with the radar products, spatial and temporal variability of rainfall rates may be lost. We assess the effects of this aggregation and issues associated with GIS processing of rainfall below.

METHODOLOGY

Ingest of Radar Data

The unprocessed reflectivity data, in its native format, is in decibels of reflectance (dbZ) for each one degree sector and 1 km range bin, and for each scan elevation. To utilize this data in a raster GIS such as GRASS, the radial data must be resampled in a gridded coordinate system. To locate the reflectivity relative to the earth's surface, the geographic coordinates, latitude and longitude, or projected coordinates such as Universal Transverse Mercator may be used. If it is important that the radar coverage appear as circles on the map produced by the GIS, then a conic projection may be more appropriate. Further corrections may be necessary due to the tilt elevation of each scan and other more minor effects associated with projections systems.

Reflectivity is extracted from 8mm archive tapes using ***r.nexrad***. This GRASS routine reads the tape and produces binary maps of reflectivity, Doppler velocity, and spectrum width at a selected nominal resolution. Once in the GRASS database, ***r.mapcalc***, is used to convert reflectivity to rainfall rate using

$$Z=300*R'$$

where reflectance, Z is in decibels and R' is the rainfall rate in mm/hr raised to a power such as 1.6. There are many operational and meteorologic factors, such as transmitter power, signal attenuation, and raindrop size distribution, that cause errors in this conversion that are beyond the scope of this paper. See Doviak and Zrnic (1984) for a description of radar applications to meteorology.

Resampling of the radial data into a gridded coordinate system is performed by taking a smaller resolution than the range bin length of 1 km. A resolution of 500 meters usually suffices in filling in the rectangular grid from the data in radial coordinates. It is important to note that the sampling resolution is a nominal resolution which is not the resolution at which the data is produced. Clearly, more information is not added by simply choosing a smaller resolution. This can be demonstrated by applying informational entropy techniques to the data over a range of resolutions (Vieux, 1993; Vieux and Farajalla, 1994; and Farajalla and Vieux, 1995).

Additional algorithms may be applied using filters that reduce the effects of outliers, hail, ground clutter, or anomalous propagation. Further, the rainfall rate should be calibrated with rain gauge amounts either in real- time or for storm totals. The rainfall rate is adjusted according to a scheme that incorporates the point information at the rain gage into the spatially distributed rainfall rate. A simple scheme is to adjust each map of rainfall rate by a factor derived from comparing storm totals estimated by rain gage and by radar. Thus, if the storm total on average from a set of rain gages is 250% greater than the storm total estimated by the radar, each radar rainfall map is multiplied by 2.5. This is one of the simplest forms of calibration. More sophisticated schemes offer improved correction at the expense of more parameters and computational effort (Nakakita et al. 1995). Selection of rain gages affects the resulting correction. RPG processing uses all rain gages under the radar umbrella. However, if a particular basin is being forecast, rain gages in or near the basin may provide a better calibration.

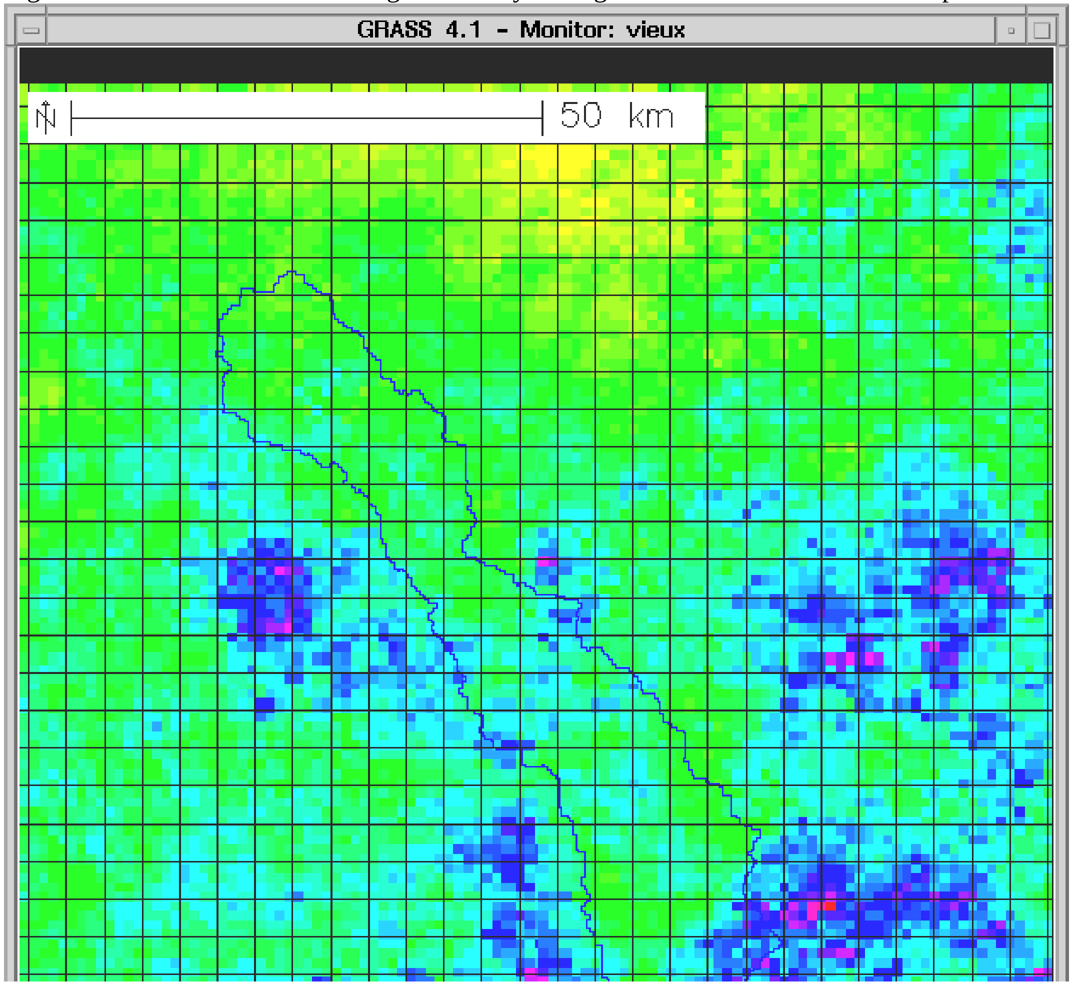
RESULTS

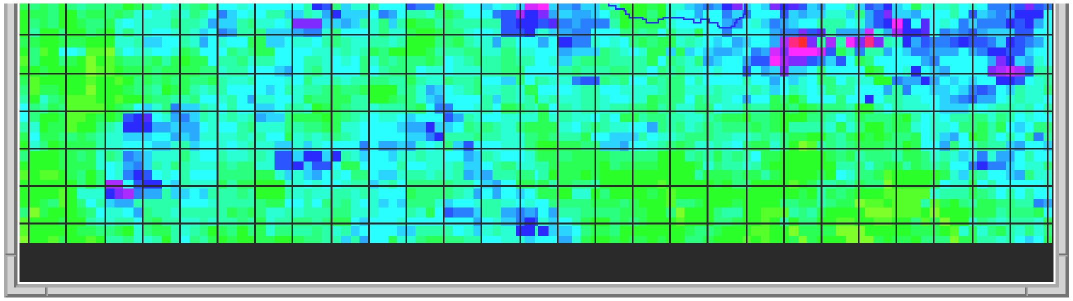
Aggregation Effects

Rainfall maps were produced for the geographic region covering the Blue basin. Maps of rainfall rate were calculated from reflectivity data obtained from the National Weather Service WSR-88D radar (NEXRAD), KFDR near Frederick, Oklahoma and KTLX, located near Oklahoma City, Oklahoma. The KTLX radar range covers the basin from about 75 km at the upper end to 175 km at the outlet. While the KFDR radar range covers the basin from nearly 180 km at the upper end to 250 km at the outlet. Five rain gages that are part of the Oklahoma Mesonetwork (Mesonet) and located in and around the basin were used to calibrate the radar rainfall estimates. Mesonet consists of 108 stations located throughout the state of Oklahoma, with at least one station located per county. These automated stations continuously monitor at 5 minute intervals rainfall and other meteorologic parameters. Data is then relayed to a central processing location every 15 minutes. Detailed rainfall rates calibrated with rain gages provide temporally and spatially variable input. In this analysis, rainfall rate, not amount, is the driving force in the storm runoff model. The effects of temporal and spatial resolution on the rainfall rate have direct influence on the hydrologic simulation.

Fig. 1 shows the storm accumulation of rainfall for an event of May 29, 1994. The grid lines are at 4 km resolution whereas the VFG radar rainfall estimates are at a nominal resolution of 0.5 km. Spatial and temporal resolution of the rainfall rates affect the simulation of storm runoff. The spatial variability of the rainfall at 500 meter resolution is lost if the maps are resampled to 4 km.

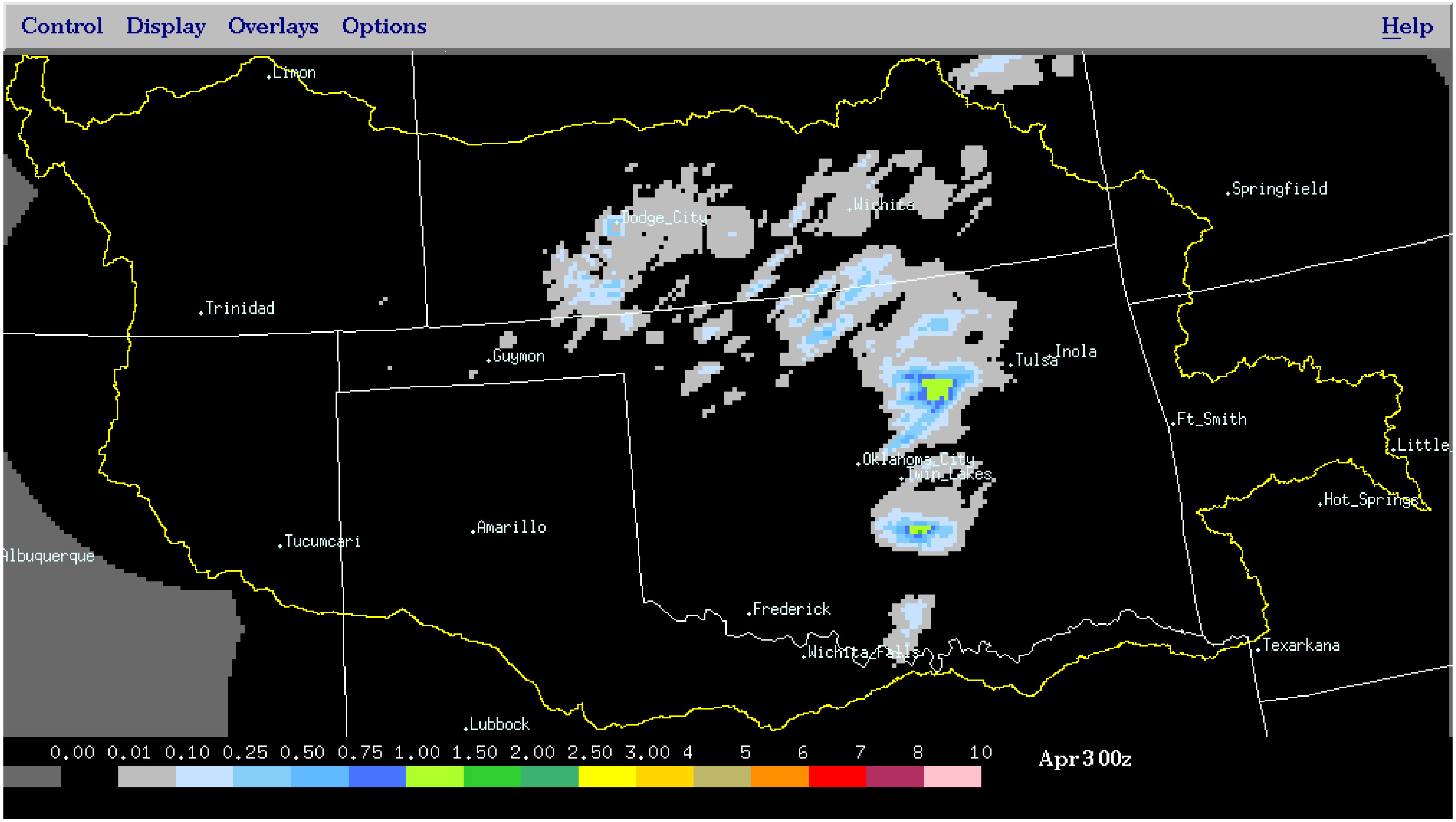
Fig. 1 Storm rainfall accumulation generated by adding each scan for the entire storm period.





The level II product of the RPG is aggregated to 4 km by the RFCs, and used in river forecasting. An example of this data set is shown in Fig. 2. The NWS RFC uses this product to forecast river stage in the Arkansas-Red Basin.

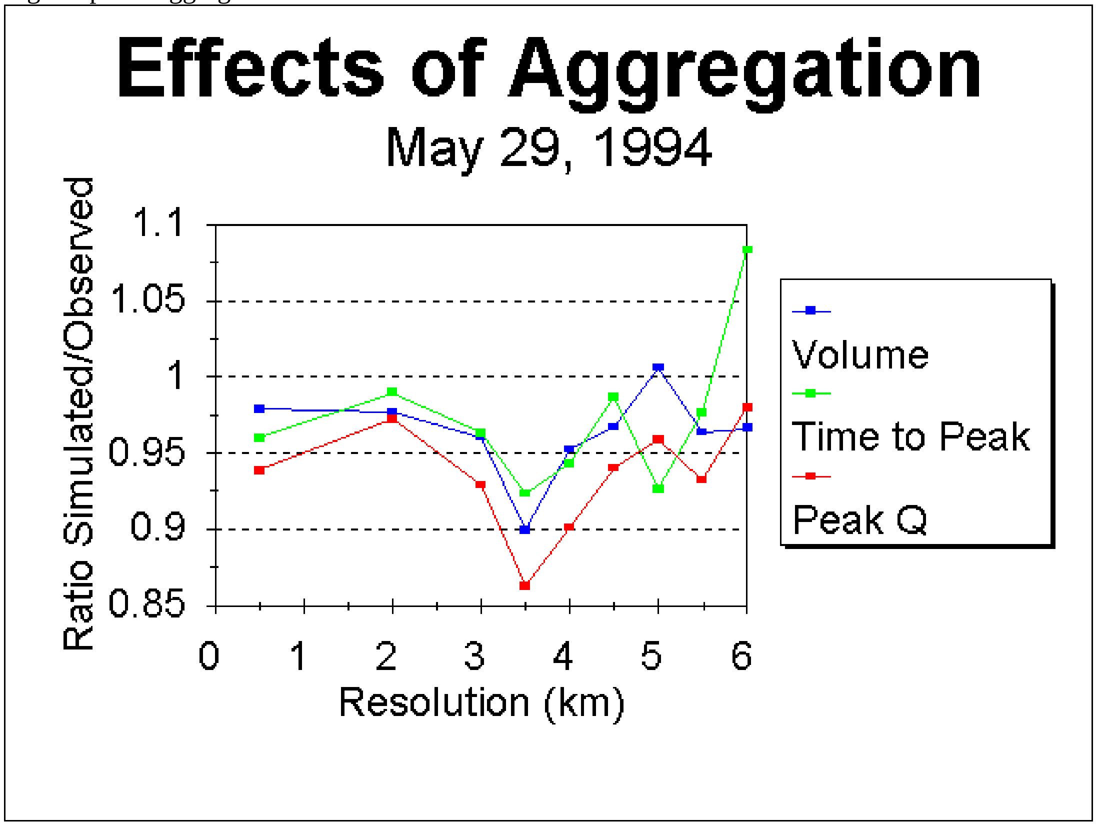
Fig. 2 NWS Stage 3 precipitation product for the Arkansas-Red River Basin.



The issue of grid cell resolution, resampling, and processing of radar estimates of rainfall is demonstrated by re-sampling the VFG 0.5 km resolution data to 6 km in 0.5 km increments. The time series of rainfall are input to ***r.water.fea*** and a hydrograph simulated for comparison with observed values. The storm that occurred on May 29, 1994 over the Blue basin is used to illustrate the effects of resolution on flood forecasting. The model was calibrated for the same storm with rainfall maps at 0.5 km resolution. While keeping the calibrated infiltration and other parameters the same, successively aggregated maps of rainfall are input to the model. The effect is seen in Fig. 3 in which three aspects are compared; volume, time to peak, and peak discharge. The ratio on the ordinate compares the simulated value to the observed value recorded by a river gaging station on the Blue river. The ratio is not one for the 0.5 km resolution due to calibration tolerances to which the model was adjusted. Progressively erratic variation results from the aggregation from 0.5 km to 6 km resolution. The variation is not severe until after 3 km. These results should be considered relative to the size of the basin simulated. The Blue River is

1200 km². Smaller basins would be more sensitive at smaller resolutions. Similar effects are observed for other storms and for aggregation in time.

Fig. 3 Spatial aggregation effects on storm runoff simulations.



Studies reported by Farajalla and Vieux (1995) revealed a similar effect of aggregation on infiltration parameters. They found that for the Blue Basin, the resolution which captured the essential variability of the infiltration parameters was ~1 to 1.2 km. Thus, from the study reported herein, the resolution that captures the essential variability for the Blue Basin for infiltration is less than that required to capture the variability in the rainfall intensities in the storm of May 29, 1994.

CONCLUSIONS

WSR-88D radar rainfall estimates are an important new source of spatial rainfall rate for distributed hydrologic models. The native resolution of the WSR-88D radar is in radial coordinates which must be resampled into a georeferenced coordinate system in the GIS. Simulations for a 1200 km² basin in south central Oklahoma reveals errors induced by the aggregation that increase in magnitude beyond 3 km for a particular storm on May 29, 1994.

A key aspect of the GIS processing is the resolution of the rainfall estimates derived from radar. The magnitude of the error and the trend with respect to resolution is dependent on the variability of the storm and size of the basin simulated. Loss of spatial and temporal variability due to aggregation of the rainfall maps is difficult to address since resolution depends on the storm morphology and the size of the basin. Neither of which is known until the storm occurs over a particular basin. Assessing the proper resolution of the rainfall estimates for use in hydrologic simulations as the data stream is being processed is needed. In this way each storm may possess a different resolution depending on the variability of the rainfall rates.

ACKNOWLEDGMENTS

Support for this research was provided by DOC-NOAA Contract NA37RJ0203, and the Cooperative Program for Operational Meteorology, Education and Training (COMET).

REFERENCES

Crum, T.D. and R.L. Alberty (1993). "The WSR-88D and the WSR-88D Operational Support Facility," *Bulletin of the American Meteorological Society*, Vol. 74, No. 9, pp. 1669-1687.

Doviak, R.J., and D.S. Zrnic, (1984), *Doppler Radar and Weather Observations,* Academic Press, Orlando, pp. 199 - 202.

Farajalla, N.S. and B.E. Vieux, (1995), "Capturing the Essential Spatial Variability in Distributed Hydrologic Modeling: Infiltration Parameters," *J. of Hydrological Processes*, Vol. 8(1), pp. 55-68.

Nakakita, E., S. Ikebuchi, K. Nakagawa, T. Sato, B.E. Vieux, and T. Takasao, "Utilization of Vertical Profile of DSD Into Building an Algorithm for Estimating Ground Rainfall Amount Using RADAR," III International Symposium on Hydrologic Applications of Weather RADAR, Aug. 20-23, 1995, Sao Paulo, Brazil.

Vieux, B.E. and N. Gaur, (1994), "Finite Element Modeling of Storm Water Runoff Using GRASS GIS," *Microcomputers in Civil Engineering*, Vol. 9:4, pp. 263-270.

Vieux, B.E. and N.S. Farajalla, (1994), "Capturing the Essential Spatial Variability in Distributed Hydrologic Modeling: Hydraulic Roughness," *J. of Hydrological Processes*, Vol. 8, pp. 221-236.

Vieux, B.E. and N.S. Farajalla, N. Gaur (1993), "Integrated GIS and Distributed Stormwater Runoff Modeling," *Proceedings of the Second International Conference on Integrated Geographic Information Systems and Environmental Modeling*, Breckenridge , Colorado Sept. 26-30, 1993, eds. M.F.

Boodchild, B.O. Parks, and L.T. Steyaert, Oxford Universiy Press, New York.

Vieux, B.E., (1993), "DEM Aggregation and Aggregation Effects on Surface Runoff Modeling," *ASCE J. of Computing in Civil Engineering*, Special Issue on Geographic Information Analysis, Vol. 7, No. 3, Jul., pp. 310-338.

About the authors...

1. Baxter E. Vieux, Ph.D., P.E. is Associate Professor in the School of Civil Engineering and Environmental Science, and the Director of the Environmental Modeling and GIS Laboratory, University of Oklahoma, 202 West Boyd Street, Room CEC334, Norman, Oklahoma, 73019, USA.

    - Homepage: http://chief.ecn.uoknor.edu/
    - Email: vieux@chief.ecn.uoknor.edu

2. Nadim S. Farajalla, Ph.D. is with Stone Environmental, 58 East State Street, Montpelier, Vermont 05602.

John E. Fels, Kris C. Matson

A Cognitively-based Approach for Hydrogeomorphic Land Classification using Digital Terrain Models

The continuous surface of the landscape is often described in discrete terms -- ridge, sideslope, flat, *etc.* -- that represent broadly recognizable cognitive types. The classification of land in this manner is useful for a variety of purposes, including the mapping of hydrogeomorphic land types associated with specific ranges of water table depth. These land types can be mapped from digital terrain models by applying decision-tree classification procedures based on pertinent geomorphometric parameters. The classification process is both interactive and interpretive, requiring repeated visualization of the resulting classification maps and adjustment of classification parameters. Moreover, it is specific to physiographic provinces having internally similar topographic and geologic characteristics affecting geomorphic and hydrologic processes.

In this study, hydrogeomorphic land types were mapped for all of North Carolina using digital terrain models with a resolution of 300 feet. Separate land type classifications were developed for each of the six major physiographic provinces of the state -- Blue Ridge, Piedmont, Sandhills, Upper Coastal Plain, Lower Coastal Plain, and Coastal Islands. The resulting classifications were then expressed in terms of expected ranges of water table depth for each land type in each province to provide an essential GIS layer for the statewide DRASTIC mapping program. These land type maps, while awaiting accuracy assessment, may also serve as a useful component of North Carolina's Corporate Geographic Database. Since these land type classifications are cognitively based, they can be broadly interpreted and might be useful in other studies requiring a regional land classification framework.

Introduction

Ground water vulnerability to contamination from the land surface has become an important area of hydrogeological research. To date, most of this research has focused on small sites of special interest using intensive sampling techniques and complex modeling procedures. There is also a need, however, for more extensive evaluations of ground water vulnerability that might help guide land planning and management at regional scales. To that end, the North Carolina Department of Environment, Health, and Natural Resources, Division of Environmental Management has undertaken the mapping of ground water vulnerability throughout the state, based on the DRASTIC system developed by the US Environmental Protection Agency (Aller *et al.* 1987).

DRASTIC is an acronym that represents seven constituent variables used in creating an overall index of ground water vulnerability to surface contamination. These variables are: average annual depth to the water table (D); net recharge (R); aquifer media (A); soil media (S); topography, expressed as land surface slope (T); impact of the vadose zone (I); and hydraulic conductivity (C). For local applications, these variables are typically determined from on-site measurements, but for regional applications they must be interpreted from existing thematic map layers. For this project, most of these variables could be reliably interpreted from geologic and soil maps but, for depth to the water table, no statewide data were available. It became necessary, then, to develop a new approach to mapping water table depth which would provide information appropriate to the DRASTIC assessment scheme, express this information at a useful level of resolution, and be applicable to very large subject areas.

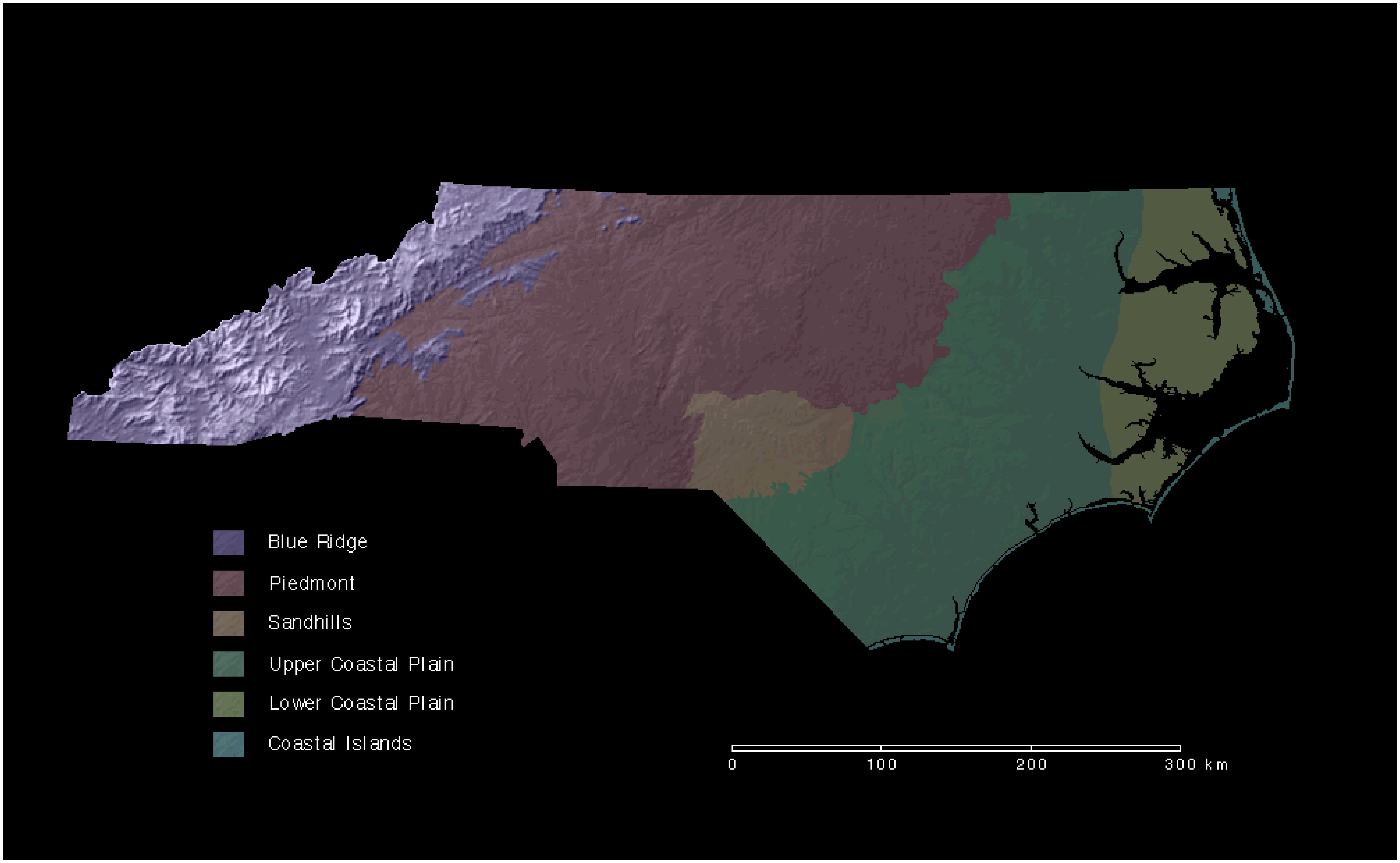
Several possible mapping approaches were considered, including deterministic modeling, statistical modeling, and landscape classification (Matson and Fels 1996). Although modeling approaches would likely provide the most accurate estimations of ground water depth, they would also require large amounts of field data -- far more than is currently available for the state. In addition, the time and costs required for modeling approaches were beyond the scope of the project. A classification approach, however, could be implemented within the scope of the project using existing data and knowledge. Hydrogeomorphic land classification based on digital terrain models could be applied to large areas at a suitable level of resolution and, since the DRASTIC system only requires ranged estimates of water table depth, would provide the information necessary for vulnerability assessment.

Hydrogeomorphic land classification divides the continuous surface of the landscape into discrete landscape elements (land types) such as ridges, sideslopes, coves, or flats, which are associated *a priori* with specific ranges of average annual water table depth. Water table depth not only varies among different land types, but among similar land types in different physiographic provinces, so it was necessary to develop a separate classification scheme for each of the major physiographic provinces of North Carolina, reflecting the particular geomorphic settings and hydrologic conditions found in each. Within each province, classification was accomplished by applying a cognitive schema differentiating land types on the basis of pertinent geomorphometric attributes. Since these land types are cognitive elements, classifications were refined through a process of repeated visualization and landscape interpretation. The present classification scheme was developed through consultation with ground water and mapping scientists from the Groundwater Section of the NC Division of Environmental Management, the Water Resource Division of the US Geological Survey, and the Design Research Laboratory at North Carolina State University using professional experience of the relationship of water table depth to different geomorphic settings among North Carolina's physiographic provinces. The classification hierarchy was limited to those land type classifications thought to be achievable using one-degree digital elevation models (the only available statewide) and classification procedures realizable within the scope of the project. It is anticipated that the classification scheme used in this first iteration of the methodology will be revised pending the results of accuracy assessment.

Methods and Results

Land type classifications were derived from 27 one-degree digital elevation models distributed by the US Geological Survey. The one-degree digital elevation models cover areas of one degree of latitude by one degree of longitude with land elevation sampled at intervals of three seconds of latitude and longitude (USGS 1990). These tiles were reprojected, using Arc/Info GRID (ESRI 1994), to the North Carolina State Plane coordinate system with a sampling interval of 300 feet. Altogether, these 27 tiles consisted of roughly 34 million model points. In addition to the gridded elevation data, corresponding tiles were also produced for land surface slope (expressed in slope percent), for landscape position, and for the major physiographic provinces of the state (Figure 1). Within each of the major physiographic provinces -- Blue Ridge, Piedmont, Sandhills, Upper Coastal Plain, Lower Coastal Plain, and Coastal Islands -- procedures for land type classification were based on various combinations of two differentiating criteria: land slope and landscape position.

Figure 1. The major physiographic provinces of North Carolina



Calculation of Landscape Position

Various quantitative methods have been developed for characterizing the morphology of land surfaces (Evans 1972, Mark 1975, Dole and Jordan 1978, Papo and Gelman 1984, Elghazali and Hassan 1986, Zevenbergen and Thorne 1987, McNab 1989, 1993, Fels 1994) and for extracting hydrologic characteristics from digital topographic models (Jenson and Domingue 1988, Skidmore 1990). However, since hydrogeomorphic classifications are not based on morphology alone but also on the position of the land surface in relation to its surroundings, a method recently developed by Fels for ecological land classification was adopted. This method yields a quantitative index of landscape position by evaluating elevation differences between a given point and other model points within a specified search radius (USFS 1995, Fels 1995). Specifically, landscape position is calculated as:

$$LPOS = \frac{\sum_{1,n} \frac{(E_n - E_0)}{d}}{n} \qquad (1)$$

where

E0 = elevation of the model point under evaluation

En = elevation of a surrounding model point

d = horizontal distance between the two model points

n = the total number of surrounding points employed in the evaluation

The value calculated is the mean of the distance-weighted elevation differences between a given point and all other model points within a specified search radius. Greater positive values indicate lower topographic positions (proximal to streams) and greater negative values indicate higher landscape positions (ridges, summits) while values approaching zero indicate mid-slope positions. Where relief is minimal within the search radius, values will also tend to approach zero.

The extent of the search area is an important consideration, since the evaluation of landscape position will be most meaningful when confined to a single landform. In principle, the radius of search should be one-half of the fractal dimension of the landscape, that is, one half of the ridge-to-stream distance in that landscape. Under such circumstances, a point located at mid-slope position will be evaluated with respect to points extending from the stream at the bottom of the slope to the ridge at the top of the slope. Average ridge-to-stream distance varies considerably among different landscapes but is fairly consistent within a particular physiographic province. Estimates of ridge-to-stream distance were obtained for the various physiographic provinces by visualizing digital terrain models within each region, measuring ridge-to-stream distance for a number of typical landforms, and taking the mean of these measurements to obtain a representative value. Average ridge-to-stream distances and search radii for the six physiographic provinces are shown in Table 1.

Table 1. Average ridge-to-stream distances and search radii for the major physiographic provinces.

Physiographic Province	Ridge-to-Stream Dist ance	Search Radius
Blue Ridge	3600 feet	6 cells

Piedmont	7200 feet	1 2 cells
Sandhills	4200 feet	7 cells
Upper Coastal Plain	6000 feet	10 cells
Lower Coastal Plain	6000 feet	10 cells
Coastal Islands	6000 feet	10 cells

Algorithms were developed to calculate landscape position index for every point in the digital terrain models, using appropriate search radii within the various physiographic provinces. Applying these algorithms to the one-degree elevation tiles produced 27 binary files of landscape position values . These were then converted to ASCII text files compatible with ARC/Info GRID.

Land Type Classification and Mapping

For each physiographic province, a classification scheme was developed to identify discrete land types and to relate these to estimated water table depth ranges for that province (Table 2). For each land type in each province, depth ranges were determined based on field experience, recorded observations, and values provided in the EPA DRASTIC Guidelines (Aller *et al.* 1987).

Within each province, land types were classified using a divisive procedure applied in successive stages (decision tree), with different criteria distinguishing land types at each stage. In the Sandhills province, for example, if landscape position at a particular point was less than a critical value, then that point was classified as a *hill* or *ridge*, otherwise if land slope exceeded another critical value that point was classified as a *sideslope*, otherwise it was classified as a *flat* (Figure 2). Classification procedures for the Coastal Plain and Coastal Islands provinces were simpler than this, while those for the Piedmont and Blue Ridge provinces were considerably more complex. Classification procedures for all six physiographic provinces are described by Fels (1995).

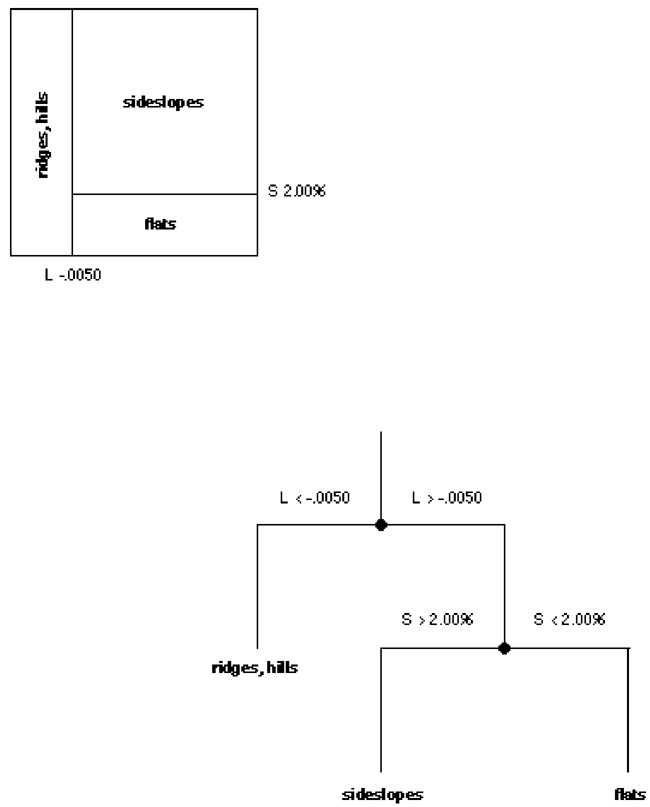
Table 2. Physiographic provinces, land types, and water table depth ranges.

Physiographic Province	**Landtype (geomorphic setting)**	**Average annual ground water depth***
Blue Ridge	*lakes, rivers, and reservoirs*	0 feet
	wetlands and wet floodplains	0-5 feet
	streamsides	0-5 feet
	floodplains and lowland flats	5-15 feet
	coves, draws, and toe slopes	5-15 feet
	sideslopes	15-30 feet
	upland flats	30-50 feet
	ridges	50-75 feet
	narrow ridges	50-75 feet

	cliffs	75-100 feet
Piedmont	*lakes, rivers, and reservoirs*	0 feet
	wetlands and wet floodplains	0-5 feet
	streamsides	0-5 feet
	floodplains and lowland flats	5-15 feet
	draws and toe slopes	5-15 feet
	sideslopes	15-30 feet
	upland flats	30-50 feet
	ridges	30-50 feet
	cliffs	50-75 feet
Sandhills	*lakes, rivers, and reservoirs*	0 feet
	wetlands and wet floodplains	0-5 feet
	flats	5-15 feet
	slopes and scarps	15-30 feet
	ridges and hills	15-30 feet
Upper Coastal Plain	*lakes, rivers, and reservoirs*	0 feet
	wetlands ,wet floodplains, and bays	0-5 feet
	flats	0-5 feet
	ridges, hills, slopes, and scarps	5-15 feet
Lower Coastal Plain	*lakes, rivers, and wetlands*	0-5 feet
	bays and wet floodplains	0-5 feet
	flats	0-5 feet
	ridges, scarps, and relict dunes	5-15 feet
Coastal Islands	*marshes and wetlands*	0 feet
	beaches, low flats, and swales	0-5 feet
	dunes, ridges, and scarps	5-15 feet

* Ranges from Aller et al. (1987).
Italics indicate hydrologic settings classified using ancillary digital geographic data.

Figure 2. Classification matrix and decision tree for the Sandhills physiographic province



S = slope value

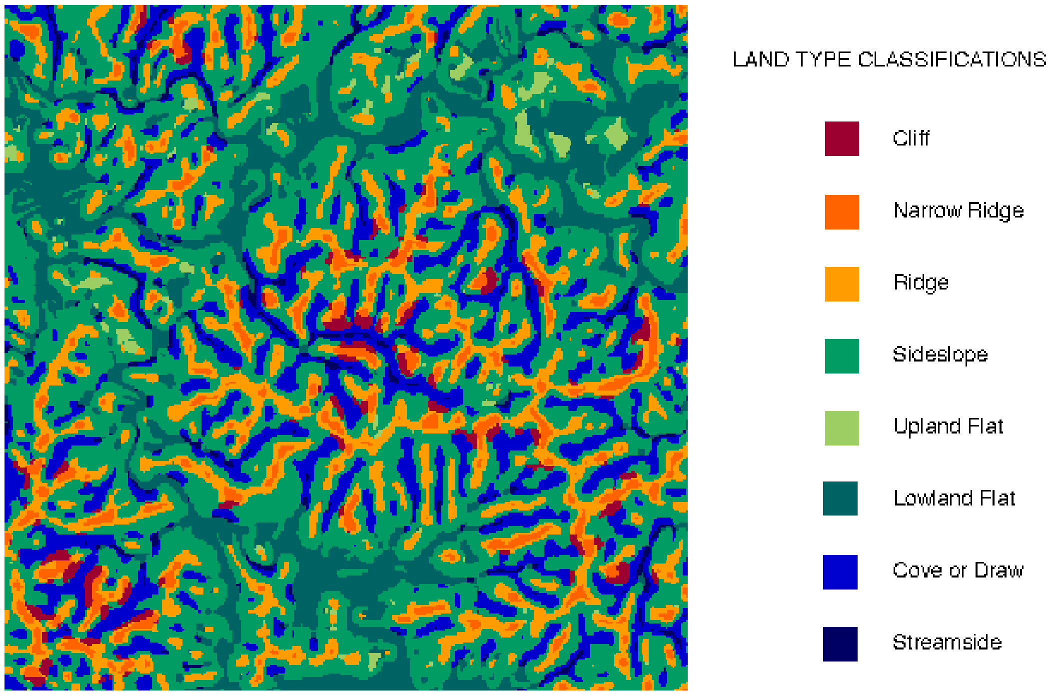
L = landscape position value

Critical values for differentiating criteria were determined by visualizing the resulting classifications, and comparing these to interpretations of land type based on familiarity with the geomorphology of the landscape and knowledge of ground water behavior in that landscape. Several representative tiles were used to determine critical values for each physiographic province. This process was repeated many times, gradually optimizing critical values and refining the classification procedures.

A special problem arose in the final stages of classification for the Blue Ridge and Piedmont provinces. This required differentiating upland flats from lowland flats based on landscape position. In these provinces, especially in the Piedmont, extensive areas of upland flat are separated from extensive areas of lowland flat by shallow to moderate slopes, forming what appear as very large terraces in the landscape. Since these 'terraces' are considerably more extensive than the typical ridge and valley landforms used to determine the search radii for landscape position in these provinces, the resulting tiles of landscape position were inappropriate for differentiating the larger upland and lowland flats. Thus, lowland classifications sometimes appeared in the center of upland flats, and vice-versa, when the dimensions of these areas were considerably larger than the specified search radius. This problem was redressed by producing a second set of landscape position tiles (termed "relative position") employing a much larger search area and providing a much more general appraisal of landscape position. That variable was employed only in the Blue Ridge and Piedmont classifications, and only to distinguish upland from lowland flats.

Land type mapping was accomplished by applying the classification procedures for each province to the 27 digital terrain models, with repeated visualization of the resulting classifications and adjustment of differentiating values. The final maps of hydrogeomorphic land types (see Figure 3) contained two-digit codes representing both physiographic provinces and land types. Again, the binary output files were converted to ASCII text files compatible with ARC/Info GRID. Since each land type was related a priori to a specific depth range, these tiles could then be reclassified to map water table depth throughout the state. To complete the statewide map of ground water depth, values for hydrologic settings (lakes, wetlands, *etc.*) were determined using digital hydrographic data from the North Carolina Corporate Geographic Database (NC CGIA 1995).

Figure 3. Classified land type map for a portion of the Blue Ridge physiographic province



Conclusions

These methods represent one solution among a spectrum of possible solutions to the mapping of water table depth for large areas. Other possible approaches include deterministic modeling, trend surface interpolation of ground water levels, regression modeling based on ground water data and larger sets of geomorphometric parameters, and landscape classification based on discriminant analysis. This approach was chosen primarily for its efficacy, given a limited timeframe and the logistic difficulty of assembling sufficient field data for direct statistical modeling. Adequate field data do exist, however, to support an accuracy assessment of the methods and assumptions embodied in this approach and, in the future, the realization and comparison of different approaches using higher resolution digital models and more complex conceptual models.

The methods employed here are clearly heuristic in nature, relying on digital media in the interactive development and testing of classification procedures. The classifications themselves are based on cognitive schemata widely employed by field scientists but, until now, not expressed in a spatially explicit context. While modeling procedures were implemented using custom software designed for that purpose and commercial GIS was used mainly to integrate modeling results, future GIS applications might better serve environmental modeling efforts by facilitating interactive modeling and visualization scenarios. Until GIS becomes a comprehensive modeling platform, environmental scientists will continue to rely on task-specific modeling software coupled with general- purpose geographic information systems.

This project has produced, in addition to its primary goal of water table mapping, a map of classified land types for the entire state of North Carolina. While this classification was designed to reflect the influence of topography on ground water depth, and was not meant to be all inclusive with respect to other fields of interest, it does provide a level of geomorphic description previously unavailable at the state level. In addition to the application described here, land type mapping might provide a useful framework for organizing ecological inventory and monitoring programs, interpreting and analyzing land use patterns, or developing land planning and management strategies. Beyond applications in ground water analyses, this information also might serve as an important component in the study and management of North Carolina's many other natural resources.

Acknowledgments

The authors wish to thank the Groundwater Section of the North Carolina Department of Environment, Health, and Natural Resources, and the United States Environmental Protection Agency, for the opportunity and funding to conduct this work. We thank Carl Bailey, Ted Mew, Arthur Mouberry, and Perry Nelson for conceptualizing and supporting efforts to characterize the hydrogeology of the shallow aquifer system in North Carolina. We also thank Ralph Heath, consultant, and Bruce Lloyd, Charles Daniel, Tim Spruill, and Gerry McMahon of the United States Geological Survey for their insights into regional water table mapping and the approaches discussed here. We are also indebted to Scott Huffman and Kai-Ping Wang of the Groundwater Section for their technical support in this mapping project. Computing hardware was provided in part by Apple Corporation, Community Affairs Division, through their EarthGrants Program.

References

Aller, L., Bennet, T., Lehr, J.H., Petty, R.J., and Hackett, G. (1987) *DRASTIC: A Standardized System for Evaluating Ground Water Pollution Potential using Hydrogeologic Settings.* Washington, DC: US Environmental Protection Agency. Document EPA/600/2-85-018.

Dole, W.E. and Jordan, N.F. (1978) Slope mapping. *The American Association of Petroleum Geologists Bulletin* 62: 2427-2440.

Elghazali, M.S. and Hassan, M.M. (1986) A simplified terrain relief classification from DEM data using finite differences. *Geo-Processing* 3: 167-178.

Fels, J.E. (1994) *Modeling and Mapping Potential Vegetation using Digital Terrain Data: Applications in the Ellicott Rock Wilderness of North Carolina, South Carolina, and Georgia.* Raleigh, NC: North Carolina State University. Ph.D. dissertation.

Fels, J.E. (1995) *Landscape Position and Classified Landtype Mapping for the Statewide DRASTIC Mapping Project.* Raleigh, NC: North Carolina State University. Technical report to the North Carolina Department of Environment, Health, and Natural Resources, Division of Environmental Management.

ESRI. (1994) *Arc-Info Version 7.* Redlands, CA: Environmental Systems Research Institute.

Evans, I.S. (1972) General geomorphometry, derivations of altitude, and descriptive statistics. *Spatial Analysis in Geomorphology*, ed. Chorley, R.J. New York: Harper and Row.

Jenson, S.K. and Domingue, J.O. (1988) Extracting topographic structure from digital elevation data for geographic information system analysis.

Photogrammetric Engineering and Remote Sensing 54: 1593-1600.

Mark, D.M. (1975) Geomorphometric parameters: a review and evaluation. *Geografiska Annaler* 3-4, Series A: 165-177.

Matson, K.C. and Fels, J.E. (1996) Approaches to automated water table mapping. *Proceedings, Third International Conference on Integrating GIS and Environmental Modeling.* Santa Barbara, CA: National Center for Geographic Information and Analysis. WWW, CD.

McNab, W.H. (1989) Terrain Shape Index: Quantifying effect of minor landforms on tree height. *Forest Science* 35: 253-258.

McNab, W.H. (1993) A topographic index to quantify the effect of mesoscale landform on site productivity. *Canadian Journal of Forest Resources* 23: 1100-1107.

NC CGIA (1995) *North Carolina Geographic Data Catalog.* Raleigh, NC: North Carolina Center for Geographic Information and Analysis.

Papo, H.B. and Gelman, E. (1984) Digital terrain models for slopes and curvatures. *Photogrammetric Engineering and Remote Sensing* 50: 695- 701.

USFS. (1995) *Ecological Classification, Mapping, and Inventory for the Chattooga River Watershed.* Atlanta, GA: US Forest Service, Southern Region, Chattooga Ecological Classification Team. Draft report.

USGS. (1990) *Digital Elevation Models, Data Users Guide 5.* Reston, VA: US Department of the Interior, US Geological Survey.

Zevenbergen, L.W. and Thorne, C.R. (1987) Quantitative analysis of land surface topography. *Earth Surface Processes and Landforms* 12: 47- 56.

John E. Fels
Research Associate Professor
Department of Landscape Architecture
North Carolina State University
Box 7701
Raleigh, NC 27695-7701
Phone: 919 / 515-7341
Email: fels@unity.ncsu.edu

Kris C. Matson
Research Associate
Department of Biological and Agricultural Engineering
North Carolina State University
Box 7625
Raleigh, NC 27695-7625
Phone: 919 / 515-6792
Email: matson@eos.ncsu.edu

Carl Gable, Harold Trease and Terry Cherry

Automated Grid Generation From Models of Complex Geologic Structure and Stratigraphy

Abstract

The construction of computational grids which accurately reflect complex geologic structure and stratigraphy for flow and transport models poses a formidable task. With an understanding of stratigraphy, material properties and boundary and initial conditions, the task of incorporating this data into a numerical model can be difficult and time consuming. Most GIS tools for representing complex geologic volumes and surfaces are not designed for producing optimal grids for flow and transport computation. We have developed a tool, GEOMESH, for generating finite element grids that maintain the geometric integrity of input volumes, surfaces, and geologic data and produce an optimal (Delaunay) tetrahedral grid that can be used for flow and transport computations. GEOMESH also satisfies the constraint that the geometric coupling coefficients of the grid are positive for all elements.

GEOMESH generates grids for two dimensional cross sections, three dimensional regional models, represents faults and fractures, and has the capability of including finer grids representing tunnels and well bores into grids. GEOMESH also permits adaptive grid refinement in three dimensions. The tools to glue, merge and insert grids together demonstrate how complex grids can be built from simpler pieces. The resulting grid can be utilized by unstructured finite element or integrated finite difference computational physics codes.

Grid Generation and GIS

Grid generation is a broad field with applications in aerodynamics, material science, biology, earth science, chemistry and physics, for example. For any problem where there is a need to represent continuous functions with discrete points, or a system of equations needs to be solved, the first step is often to construct a discrete tessellation of the region. In applications where complex geometries do not need to be represented, regular orthogonal grids are simply constructed. When it is important to represent complex geometries, multiple materials and properties, and optimize the shape of the elements used for tessellation, automatic grid generation tools are necessary.

Grid generation tools must provide the means to build and represent volumes from surfaces, distribute nodes within a volume, define a connectivity of nodes that is compatible with the computational tools being utilized for modeling, and provide a means of assigning initial and boundary conditions. In addition, grid generation must insure the quality of the grid to insure the accuracy and stability of the physics being modeled.

There are a wide variety of geologic applications where accurate representation of complex engineering systems (wells, tunnels, reservoirs) and geologic structure and stratigraphy (layering, folding, domes, faulting) is critical to producing accurate numerical models of fluid flow and mass transport. Oil and gas reservoir production, groundwater resource development, hazardous waste site characterization and remediation, and nuclear waste disposal in a geologic repository are examples of the areas where modeling must be used to predict the long term behavior of a system. In all the systems, grid generation is a key link between the geoscientific information systems (GSIS, Turner, 1991) and the numerical models. They must capture complex geometry and insure the computational grid is optimized to produce accurate and stable solutions.

Issues dealing with the general approach and methods of integrating GIS and various process modeling studies are presented in Fisher (1993) and Moore (1993). Although many of the GSIS modeling methods and tools (Fisher, 1993) claim to be developed as solid geometry modeling tools they often lack open architecture or the ability to export the models they create. Consequently, their use as a preprocessor for physics packages is limited.

This work identifies the data needs for a particular grid generation tool, GEOMESH (Gable, 1995), discusses data that is required for GSIS to be useful as a preprocessor for grid generation and physical modeling, and presents examples of grid generation projects utilizing GEOMESH. This discussion concentrates on integrating GSIS data into deterministic numerical models and discusses issues in producing optimized grids which accurately maintain geometries defined in GSIS. Deterministic 3D geometry modeling assumes that the goal is to accurately represent the GSIS model. It does not insure the quality of input data or how errors in input data affect the GSIS model.

Structured and Unstructured Grids

Grid generation is far more than the tessellation of space with polyhedra. There are two major classes of grids, structured and unstructured. In the simplest case, a structured grid is a logically rectangular array of points with a well-defined and regular relationship of each point with its neighbor. For example, a typical 2D finite difference grid (Figure 1) can represent a function f(i,j) in a regular pattern where the neighbor to the right is f(i+1,j) and the neighbor above is f(i,j+1). A 2D rectangular grid will have four connections to any interior node, however they are not necessarily orthogonal. Irregular structured grids can be thought of as regular orthogonal grids but with all of their connections made of rubber bands so the grid can be stretched to a complicated shape as long as none of the connections are broken and none of the grid elements are turned inside out. This type of gridding scheme has many limitations, primarily with the complexity of geometries that can be represented. Numerical schemes for solving equations on these grid often have problems with accuracy, as the stretched grid becomes nonorthogonal.

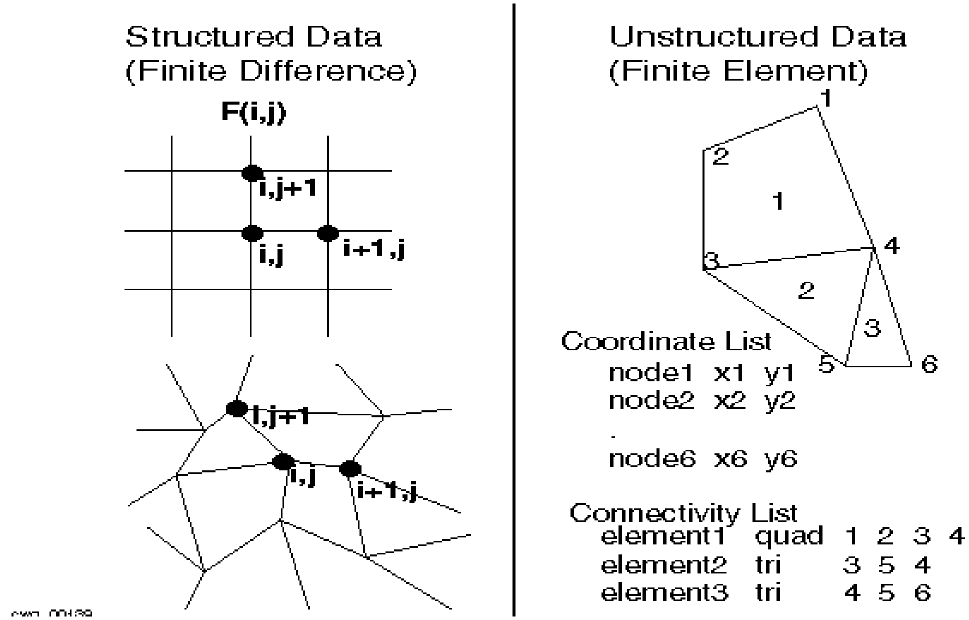


Figure 1. Nodes and connections in a structured grid have a fixed relationship to one another. Unstructured grids need explicit statements of the connectivity between nodes.

Unstructured grids, often used in finite element computations, (Figure 1) offer far more flexibility in representing complex geometries. Grids can be constructed from a variety of elements (triangles, quadrilaterals, tetrahedra, hexahedra). This flexibility comes at a higher cost. The spacial relationship of one element to another must be explicitly stated since there is no implied or logical relationship of one element to its neighbor. The number of connections a particular node has can vary greatly.

GEOMESH Grid Generation

GEOMESH is a software tool for automatically producing unstructured finite element grids tuned to the special needs of geologic and geo-engineering applications. The code provides for 2D and 3D grids and includes elements that are triangles, quadrilateral, hexahedral and tetrahedral. Capabilities have been expanded to improve grid generation for complex geometries and for providing more choices in designing a grid. At the same time, old routines have been tested and solutions found for degenerate cases. The design of the code is modular, allowing for flexibility and consistency.

The core functions of GEOMESH utilize the X3D grid generation package developed at Los Alamos National Laboratory. GEOMESH developed out of a need for accurate and automated grid generation for 3D modeling of subsurface porous flow and transport. Since the grids represent the geology being modeled, the accuracy of the grid directly affects the accuracy of the model. It was also found that grid generation was tedious, time consuming, and prone to errors, especially for models with complex structures such as faults and stratigraphy such as pinch outs and layer truncations.

Automated grid generation algorithms not only streamline grid construction, they offer more flexibility and consistency. As input constraints change, and as better data sets become available, these data are easily incorporated into the computational mesh. Automated grid generation can also be used to produce coarsened grids for preliminary calculations and refined grids for final, high resolution calculations. The grids produced by GEOMESH are widely applicable and can be used by any numerical algorithm that can utilize unstructured grids. They are not specific to any particular computational code

Defining a geometry is the first step in grid generation and is where the grid generation tool must interface with GSIS systems. The GEOMESH program can input geometry definitions in terms of their bounding surfaces. The bounding surfaces may be defined by triangular irregular networks (TINS), multi-dimensional analytic function representations such as nonuniform rational B-splines (NURBS), or by regular grids f(i,j) where f is the grid elevation and (i,j) are regular spaced x and y coordinates. Regardless of the means of defining surfaces, it is important that the geometry defined by the surfaces is unambiguous. Surfaces must form a closed volume, the normal to each surface must be consistently defined, the line defined by the intersection of two planes must be unique and there cannot be gaps where the interface of one surface joins another surface.

An alternative to surfaces are representations that already define volumes. For example, there are some GSIS packages that produce a mesh as output. However, the mesh may not be usable for computations. In this case the grid generation tool must fix the problems to allow computations while maintaining the geometry of the GSIS model. This utilizes grid generation tools for insuring grid quality.

GEOMESH grid generation uses three criteria to insure grid quality. They are: (1) the final grid preserves the input geometry model, (2) the grid is Delaunay (Figure 2) and (3) all coupling coefficients are positive. It is beyond the scope of this work to give the mathematical details of Delaunay triangulation. However, an intuitive definition of a Delaunay triangulation is that the circumscribed circle of any triangle contain the three points of the triangle and no other points. The extension to three dimensions is that the circumscribed sphere of any tetrahedra contains the four points of the tetrahedra and no other points. There are special cases and degeneracies when the above description is not absolutely correct, but these cases are not important to this discussion. The third constraint, that the coupling coefficients related to grid geometry produce a semi-positive definite matrix (Trease and Dean, 1990, Gable et al. 1995), insures that the solutions to nonlinear flow and transport will be stable and accurate. In addition, there are no negative transmissibilities when solving porous flow problems.

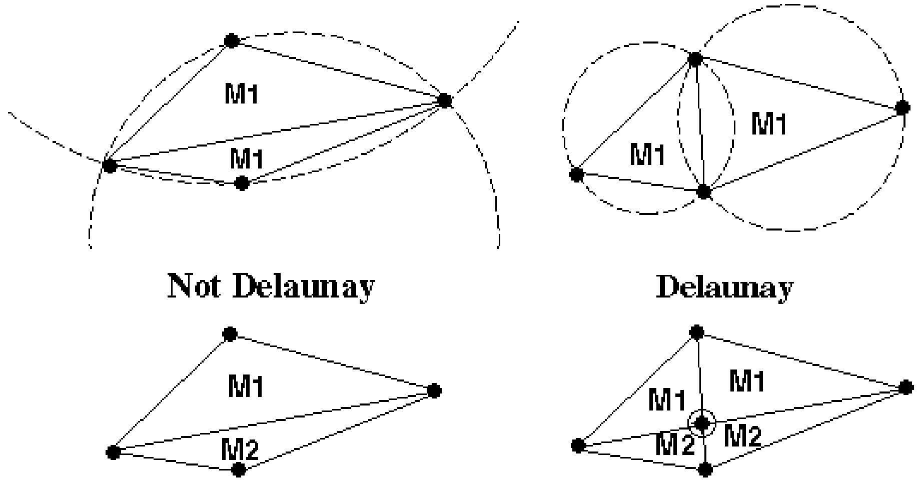


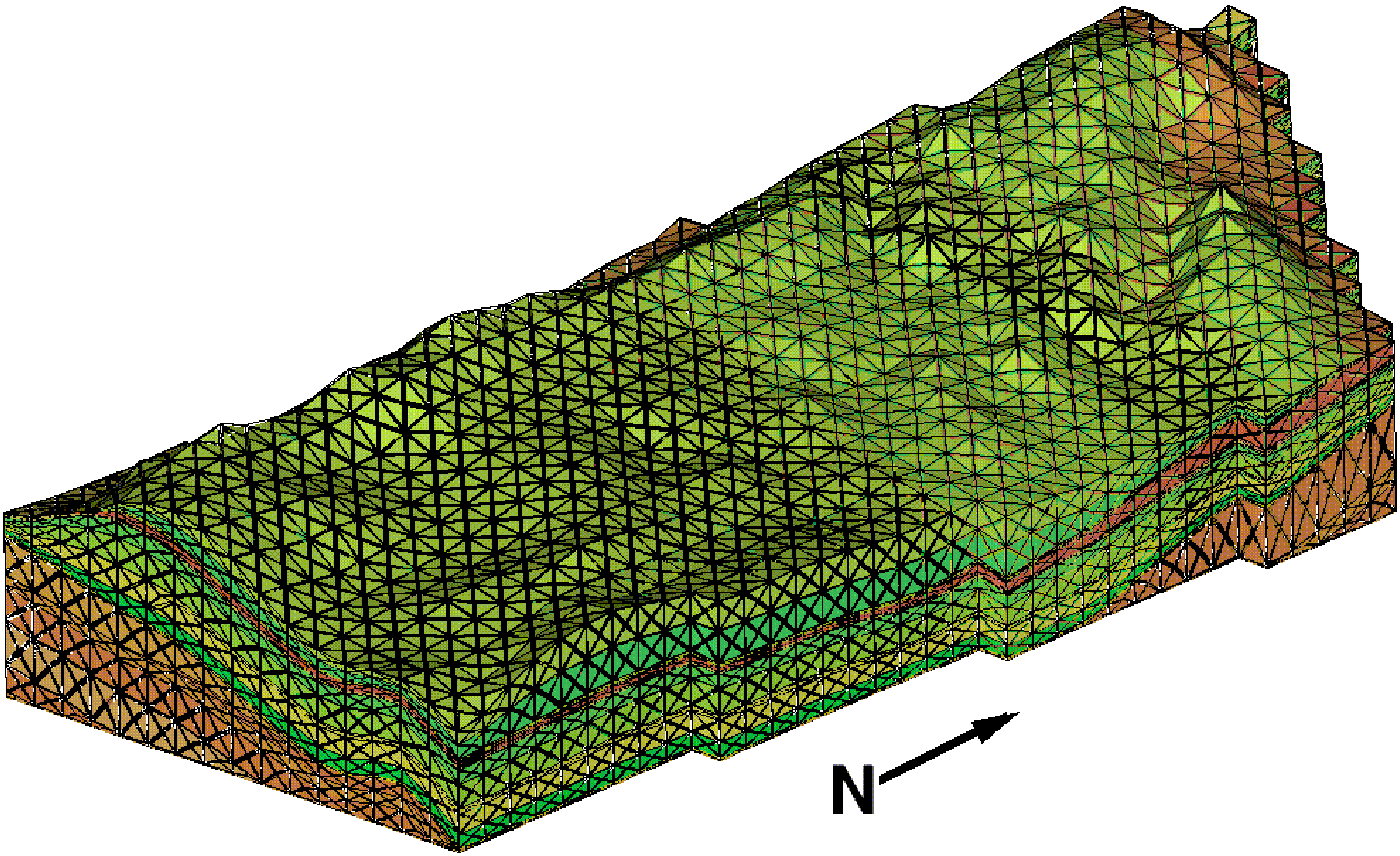
Figure 2. Example of four points connected to form a Delaunay triangulation. In the upper figure both triangles are the same material so the triangulation is made Delaunay by flipping the connection. In the lower figure, the triangles on the left are different materials, so the triangulation is made Delaunay by adding a point on the interface and making four triangle from the original two. (Figure from Gable et al. 1995.)

Many of the problems associated with importing data into grid generation tools results from insufficient GSIS output to completely characterize the solid geometry model. This may be a consequence of ambiguities in the output geometry model. Some GSIS packages produce beautiful graphical representations of the geometry model, but do not provide utilities or an open architecture for exporting the solid geometry model.

Examples

The first example, (Figure 3) shows a 3D grid produced from a GSIS 3D geometry model. The GSIS model is composed of hexahedral (8 node, 6 sided elements). However there are many degenerate elements that must be eliminated. For example, when a layer pinches out the GSIS model retaines zero volume elements as part of the output. While this produces a correct representation of the geometry, physics codes cannot compute on zero volume elements. Furthermore, the GSIS output produces elements without regard element shape or grid quality.

In order to produce a grid for computation from the GSIS output the following steps are taken: (1) extract the 3D geometry model, (2) convert the hexahedral elements to tetrahedral elements by subdividing, (3) eliminate the zero volume and degenerate elements, (4) reconnect the grid to insure a Delaunay grid while never disrupting material interfaces, (5) subdivide elements to insure positive coupling coefficients and(6) output node coordinates, connectivity, coupling coefficients and a list of boundary and material nodes. This grid is then used for computation of unsaturated porous flow and contaminant transport (Robinson et al. 1995).


N

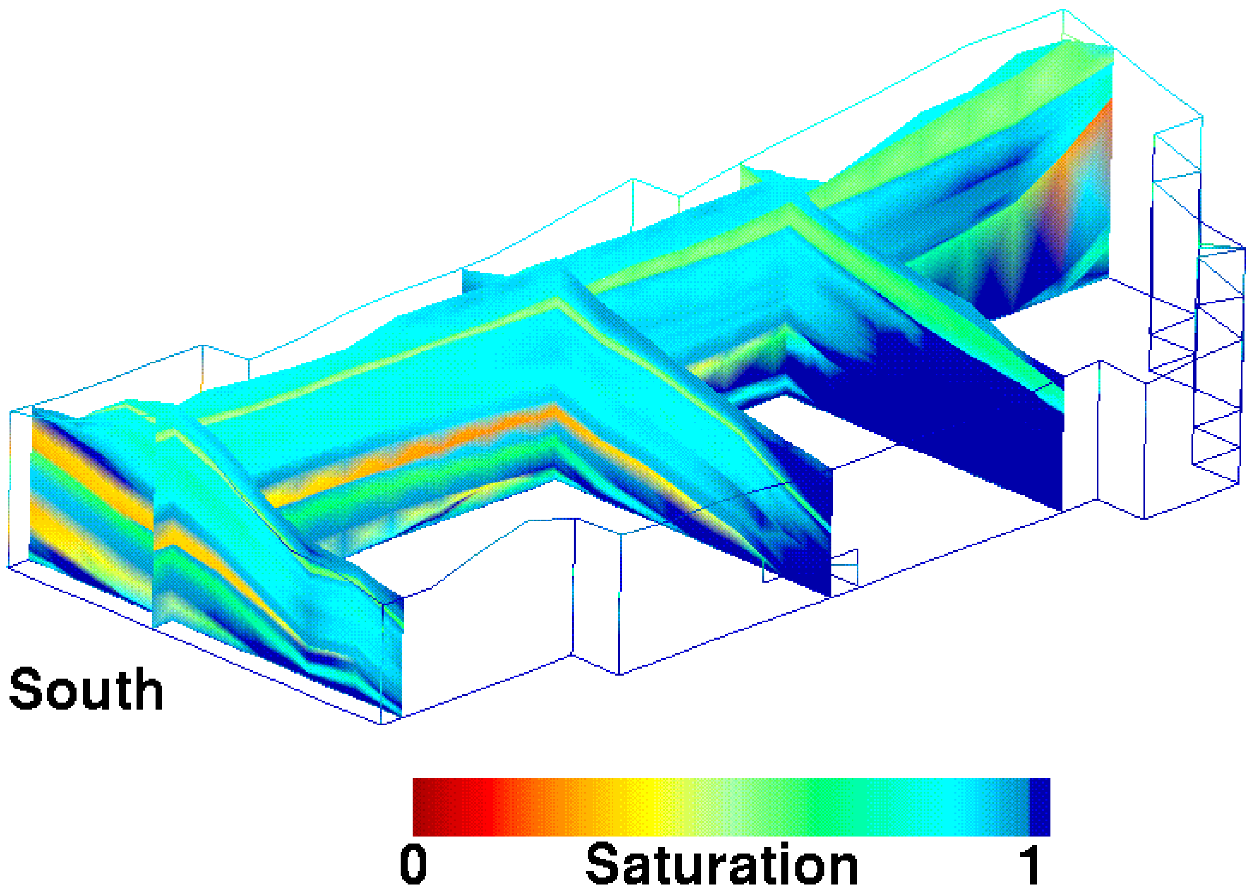

South
0
Saturation
1

Figure 3. **3D grid produced from 3D GSIS geometry model. The model is used to calculate saturation due to surface infiltration. Steady state saturation is shown in the lower figure. (Figure from Robinson et al. 1995.)**

The second example, Figure 4, is a 2D cross section extracted from the 3D model shown in Figure 3. All of the material interfaces in the geometry model are maintained, and the computational grid quality has been insured by producing a Delaunay grid with all positive coupling coefficients.

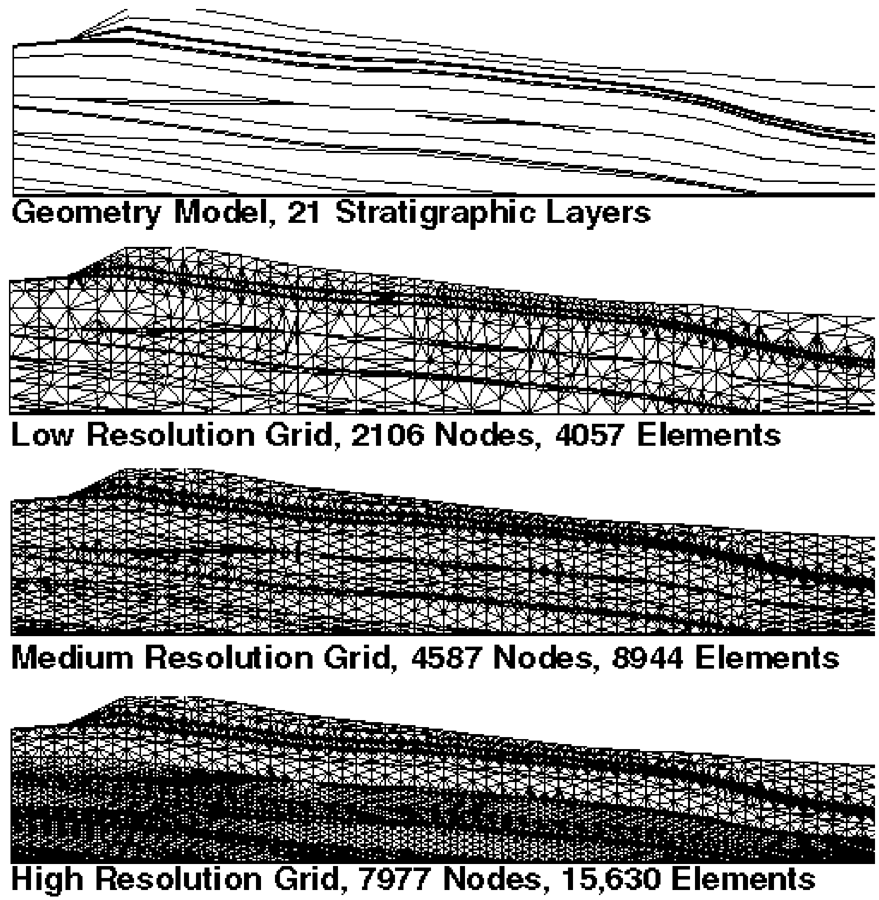


Figure 4. Three grids produced from the same geometry model of stratigraphy. Low resolution grids are used for initial calculations and high resolution is used for final high accuracy calculations. All the grids maintain material interfaces, are Delaunay and have all positive coupling coefficients. (Figure from Robinson et al. 1995)

These cross sections have a vertical extent of approximately 700 meters with very thin layers of a few meters. The automatic grid quality algorithms made it simple to produce these grids with very large differences in length scale from one layer to another.

A final example demonstrate the ability to add or merge one grid with another. This provides a powerful tool for building complex grids from simpler building blocks. In this example (Figure 5), a mesh representing a curved well is inserted into a background grid. This mesh merge utility can be used in a wide variety of other applications. For example, instead of using grid refinement to produce a high resolution grid, one can produce two grids, a course grid that covers a large volume and a high resolution grid that covers a smaller volume and merge them together to produce a single grid with high resolution in a specific region.

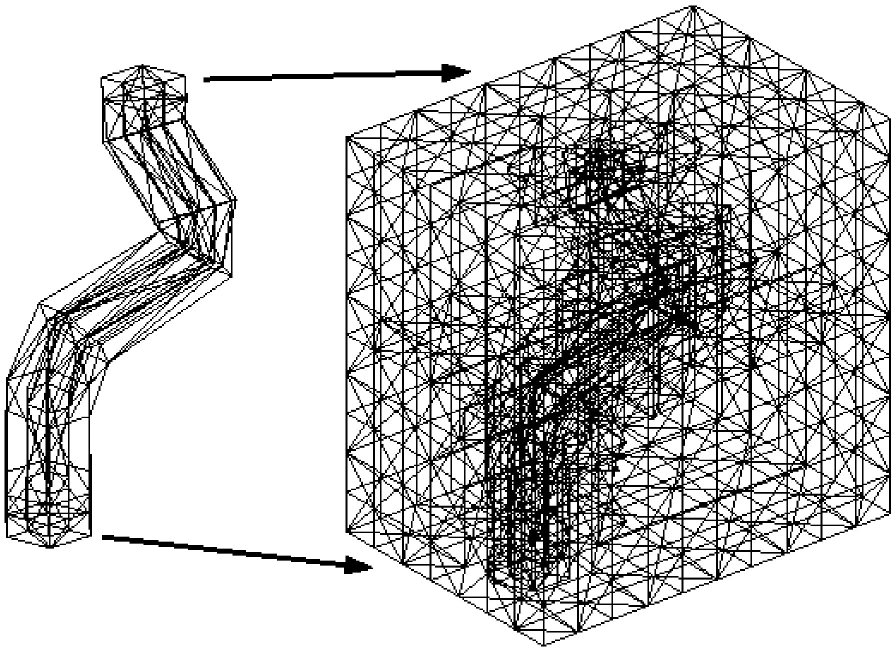


Figure 5. Complex grids can be formed by merging together simpler pieces. In this example a curved well is merged with a uniform background grid. (Figure from Gable et al. 1995.)

Other features not demonstrated by these examples are interpolation functions and grid smoothing. Interpolation of node attributes (i.e. temperature, pressure) from one grid to another is useful when applying grid refinement algorithms so that the refined grid also has estimates of attribute values. Grid smoothing algorithms allow grid optimization by letting nodes move rather than by changing connections. This method preserves material interfaces fixing the position of interface nodes while allowing nodes in the interior to move until grid quality constraints are achieved. This method can also be used to increase grid resolution where necessary by allowing the grid to migrate to zones where finer gridding is needed.

Future Outlook

There are a number of GSIS software packages available whose market is mainly oil and gas reservoir modeling. However, most of these packages suffer from a lack an open architecture. These tools often provide means of building complex representations of complex geology, however they do not provide open, accurate and flexible means of exporting 3D geometry models once they are created. Software packages that produce impressive 3D visualizations are a dead end for getting the information into physics modeling packages. This severely limits their utility as a tool in a fully integrated approach to computational physics modeling.

As tools for various aspects of an integrated modeling approach mature and become more sophisticated, the problems that the user often faces are issues of data compatibility. Well logging software may not talk to the GSIS software. The GSIS software may not interface with the grid generation tools. The grids produced are not tuned to the special needs of a physics code. The results of physics modeling cannot be easily used in reinterpretation and modification of the GSIS model. As each of the separate pieces mature, more effort will need to be spent creating seamless connections amongst them. The ability to incorporate and utilize a wide variety of data into modeling and interpretation will have a strong impact towards achieving the goal of fully integrated modeling. Automated grid generation is one piece of this puzzle.

References

Abdou, M. K., H. D. Pham and A. S. Al-Aqeeli, (1993), Impact of grid selection of reservoir simulation, Journal of Petroleum Technology, 45(7), pp 664-669.

Fisher, T., (1993), Use of 3D geographic information systems in hazardous waste site investigations, Environmental modeling with GIS, ed. Michael F. Goodchild, Bradley O. Parks and Louis T. Steyaert, Oxford University Press, pp 238-247.

Gable, C. W., T. A. Cherry and H. E. Trease and G. A. Zyvoloski, (1995), GEOMESH grid generation, LA-UR-95-4143.

Gable, C.W., George Zyvoloski, (1994), Site Scale Modeling of Radionuclide Transport At Yucca Mountain, NV: Grid Generation and Reactive Tracers, LA-UR-94-1041.

Khamayseh, A. and Andrew Kuprat, (1995), Anisotropic Smoothing and Solution Adaption for Unstructured Grids, LA-UR-95-2205, International Journal for Numerical Methods in Engineering, (submitted).

Moore, I. D., A. K. Turner, J. P. Wilson, S. K. Jenson and L. E. Band (1993), GIS and Land-Surface-Subsurface Process Modeling, Environmental modeling with GIS, ed. Michael F. Goodchild, Bradley O. Parks and Louis T. Steyaert, (Oxford University Press, pp 196-230.

Robinson, B. A., A. V. Wolfsberg, G. A. Zyvoloski and C. W. Gable, (1995), An unsaturated zone flow and transport model of Yucca Mountain, in press.

Trease, H. E. and S. H. Dean, (1990), Thermal Diffusion in the X-7 Three-Dimensional Code, Proceeding of the Next Free-Lagrange Conference, Jackson Lake Lodge, Wyoming, June 3-7, Springer-Verlag Press, Vol. 395, pp. 193-202.

Turner, A. K., (ed.) (1991), Three Dimensional Modeling With Geoscientific Information Systems, Nato ASI Series C: Mathematical and Physical Sciences, Vol. 354, Dordrecht: Kluwer Academic Publishers.

Author Information

Carl W. Gable

mail MS F665, Los Alamos National Laboratory, Los Alamos NM 87545

email gable@lanl.gov

voice 505-665-3533

fax 505-665-3687

Last Modified: 03:15pm MST, January 05, 1996

Brian Lees

IMPROVING THE SPATIAL EXTENSION OF POINT DATA BY CHANGING THE DATA MODEL

Abstract

The most common form of delivery, and use, of vegetation and soils information remains the mapped thematic, choropleth, form. The pre-processing of data to suit this data structure perpetuates the use of an inappropriate data model and places an upper limit on the accuracy of spatial extension of point data by most predictive modelling techniques. In many cases a continuum of change is being represented as a series of overlapping gaussians. This leads inevitably to the generation of errors of omission and commission. These are artefacts resulting from the choice of an inappropriate data model. This is not a new insight, Sinton dealt with it in 1978, but it remains a perennial one. The extensive work on the Kioloa data set to develop and test new predictive modelling tools was constrained to use standard (forest) industry data as input, and forest types as output. This resulted in an upper limit to predictive accuracy, for all the models tested, of about 65%. Those themes which were more appropriately represented by this data structure (land/sea discrimination) could be predicted with up to 99% accuracy. Using the same techniques, data points and variables, but changing the data model, it is possible to achieve a considerable increase in predictive accuracy. Expressing species distribution as a fuzzy membership of the tallest stratum (in a forest), midstratum or understorey, and lower stratum or ground layer allows the prediction of a series of data layers which represent surfaces of spatially varying fuzzy memberships. Further, the use of a simple neural net configuration enables both fuzzy membership and the probability of this membership to be estimated. This makes it possible to track error through subsequent uses of the modelled estimates. Methods of comparing the two methods are still under investigation, but it is clear that the change in data model results in significant gains in accuracy.

Introduction

Current approaches to the analysis of spatial data tend to deal with the spatial characteristics of the data as another attribute in an n dimensional, analytical space. The most common analysis is some form of cluster analysis. It is often an illogical framework for analysis. This has been discussed elsewhere (Lees, 1994, 1995; Aspinall and Lees, 1994). Topological relationships which exist in geographic space are lost and illogical relationships are created. Similarly, topological relationships which exist in, say, environmental data space, and form the basis for environmental domain analysis, are disrupted. This suggests that the practice of creating a single, n-dimensional, space to facilitate the analysis of spatial data destroys some of the most important characteristics of the data prior to analysis. In the following section I wish to review the previous discussion of this view of data and then move on to consider the implications of this for predictive accuracy.

Operational Data Spaces and Appropriate Data Models

Spatial data exists in a number of discrete domains (Lees, 1994, 1995; Aspinall and Lees, 1994) . In each of these there exist topological relationships, but these relationships vary from domain to domain. We are most familiar with spatial data existing in a geographic space defined by latitude, longitude and elevation. Movement from point to point in this space is a vector. It is not possible to move from one point to another without transiting intermediate points. Each point is unique.

In the other, conceptual, domains or data spaces topological relationships are different. These data spaces can be spectral space, environmental data space, even socio-economic data space. The fundamental, and shared, characteristic of these spaces is that movement through the space has a logical meaning. Spectral space, for example, forms the basis for most analysis of remotely sensed data. Proximity suggests similar colour. Trajectories of reflectance values for developing crops on different soils form the basis for the common Kauth-Thomas, or Tasseled Cap, transformation (1976). Trajectories in spectral space form the basis for sub-pixel modelling of vegetation structure (Jupp et al, 1986). In these analyses vectors represent changes in the reflectance at a point, through time. No motion in geographic space is envisioned. A large number of points in geographic space can occupy a single location in spectral space. The converse is not true.

In environmental data space, the basis for environmental domain analysis (Mackey et al., 1988), topological relationships are linked directly to environmental gradients. Vectors in this space drive the continuum of change in vegetation composition observed in nature. The parameters typically used in environmental domain analyses tend to act at the species and not community level. The conflict in ecological literature between those who favour a community view of vegetation and those who view it as a continuum lies squarely on the fact that community is a spatial concept in geographic space, whilst the continuum is a spatial concept in environmental data space (Austin and Smith, 1989). These are fundamentally different in the way they can be analysed. In geographic space one can move from one point to another along a vector. This same motion in environmental data space may result in no motion, if the environments along this vector in geographic space are the same, or a jump from point to point if say, a soil boundary is crossed. As before, a large number of points in geographic space can occupy a single location in environmental data space and, once again, the converse is not true.

This particular dichotemy, between representation of vegetation distribution in geographic space and environmental data space, is a dichotomy between data models. The `mapping' school reduce observations of vegetation to a series of vegetation classes, even forest types. In many ecosystems the class boundaries are cultural (statistical) artefacts. Slight changes in contribution to the canopy can lead to a change in class. In some cases, there is often more variation within the class than between classes. Nevertheless, the fundamental structure of choropleth mapping requires this reduction of variance to permit the mapping of polygons. This mismatch between data and data structure, excusable in the days where choropleth mapping was the only means of representation, has been carried forward to the present. The often uncritical use of an incorrect model to represent a particular spatial distribution is a persistent source of error in geography (James, 1967).

If we take, as an example, vegetation observations from a typical, undisturbed, eucalypt forest in hilly terrain these can be plotted in geographic space, environmental data space and spectral space. Examining these field observations in environmental data space shows very clearly that much of the variance in the data forms a continuum. Mapping this continuum across to geographic space `shatters' this continuum into facets. These do look like polygons until one examines their internal behaviour. Each facet retains some of the continuous variation which formed the continuum in environmental data space.

`Facets' are not a recognised data structure, but the assumption that, because they appear similar to polygons in geographic space, the latter is an appropriate data model for vegetation, is a major source of error in vegetation and soil mapping. Because choropleth mapping requires the observed variability of data to be reduced, pre-classification is required. The mapping errors are compounded because this pre-classification of the vegetation into communities is carried out in taxonomic data space. The *implied* distribution is now a series of overlapping gaussians. Attempting to represent a continuum as a series of overlapping gausians means that the predictive accuracy of the analysis drops markedly because of decision errors of omission and commission (Sinton, 1977).

This reduction in accuracy is directly related to the use of an inappropriate data model. It would be more appropriate to model this phenomenon in environmental data space or without pre-classification of the observations (Payne, Stockwell and Davey, 1994).

Mapping Vegetation Attributes Without Pre-classification

We used the Kioloa data set (Lees and Ritman, 1991), this covers an area of complex terrain with elevations ranging from sea level to 285m. Land cover varies from rainforest, highly disturbed forest, heath to cleared grassland. Extensive analyses of this dataset using the ground data pre-classed into forest types typically give predictions of `Sea' at better than 95%, `Paddock' at better than 80% and the other, forest types at accuracies between 45%-65%. Treating the continuum of change within the forest as a set of overlapping gaussians, or classes, means that better results are impossible using classed data. However, if one deals with the point observations of vegetation without pre-classification, then another form of mapping must be used. Because we are now dealing with digital information, we are no longer constrained to produce a single `map'. At any point We can now store and retrieve a considerable amount of information about any point, either in geographic, environmental or spectral space. If we retain geographic space as the most convenient operational data space for users of predictive modelling, we are also selecting the domain with least database complexity where each point exists in only one location in each of the other domains. We can then model the spatial extension of each relevant attribute of each entity observed in the ground truth plots. It is possible to add an estimate of the error at each location resulting from this.

Using the Lees and Ritman (1991) dataset the original observations can be recoded as DBH, stem densities, biomass, or canopy contribution. For this example, the last was chosen. As this is a genuinely fuzzy phenomenon, it was coded as fuzzy membership of the canopy, species by species. A simple Back Propagation neural net was set up for each species. That developed for Eucalyptus maculata, spotted gum, is described. In order to provide an ongoing comparison of methods, the input layer was the same as that described in Fitzgerald and Lees (1993, 1994), and used the same datasets as Lees and Ritman (1991). A 9/10/10 structure was used with one hidden layer of ten nodes and an output layer of ten nodes. No spatial or temporal context was used.

The network used the Delta rule with a sigmoid transfer function. Learning rates were set at 0.9 initially (0-5000), then reduced steadily (0.3 for 5000-10,000; 0.2 for 10,000 - 150,000). Initial runs suggested that the momentum term needed to be set rather higher than usual (0.6 for 0-20,000; 0.05 for 20,000-150,000).

Each output node represented a range of fuzzy memberships (0-0.1; 0.1-0.2 and so on) (fig 1). The highest number in the output range was taken to indicate the membership of that cell and the whole output range for each cell was treated as a distribution and the probability of the membership was calculated. So, for each species, it is possible to estimate its fuzzy membership of the canopy (figs 2 & 3) and the probability of that estimate (fig 4).

fuzzy

membership cell 0/row 0 cell 1/row 0

0-0.15	1.086	0.908	0.749	0.429	0.414	0.472	0.495
0.2	0.086	0.152	0.2	0.268	0.287	0.274	0.271
0.3	0.013	0.008	0.01	0.009	0.008	0.01	0.01
0.4	-0.072	-0.004	0.034	0.087	0.078	0.115	0.108
0.5	-0.001	0.011	0.029	0.058	0.058	0.051	0.05
0.6	-0.013	-0.013	-0.01	-0.009	-0.015	-0.014	-0.014
0.7	0.006	0.035	0.073	0.151	0.162	0.106	0.104
0.8	-0.052	-0.044	-0.025	0.009	0	-0.011	-0.014

Figure 1: Cell by cell output from the NN for row 0, cells 0-6.

0.9 0 0 0 0 0 0 0

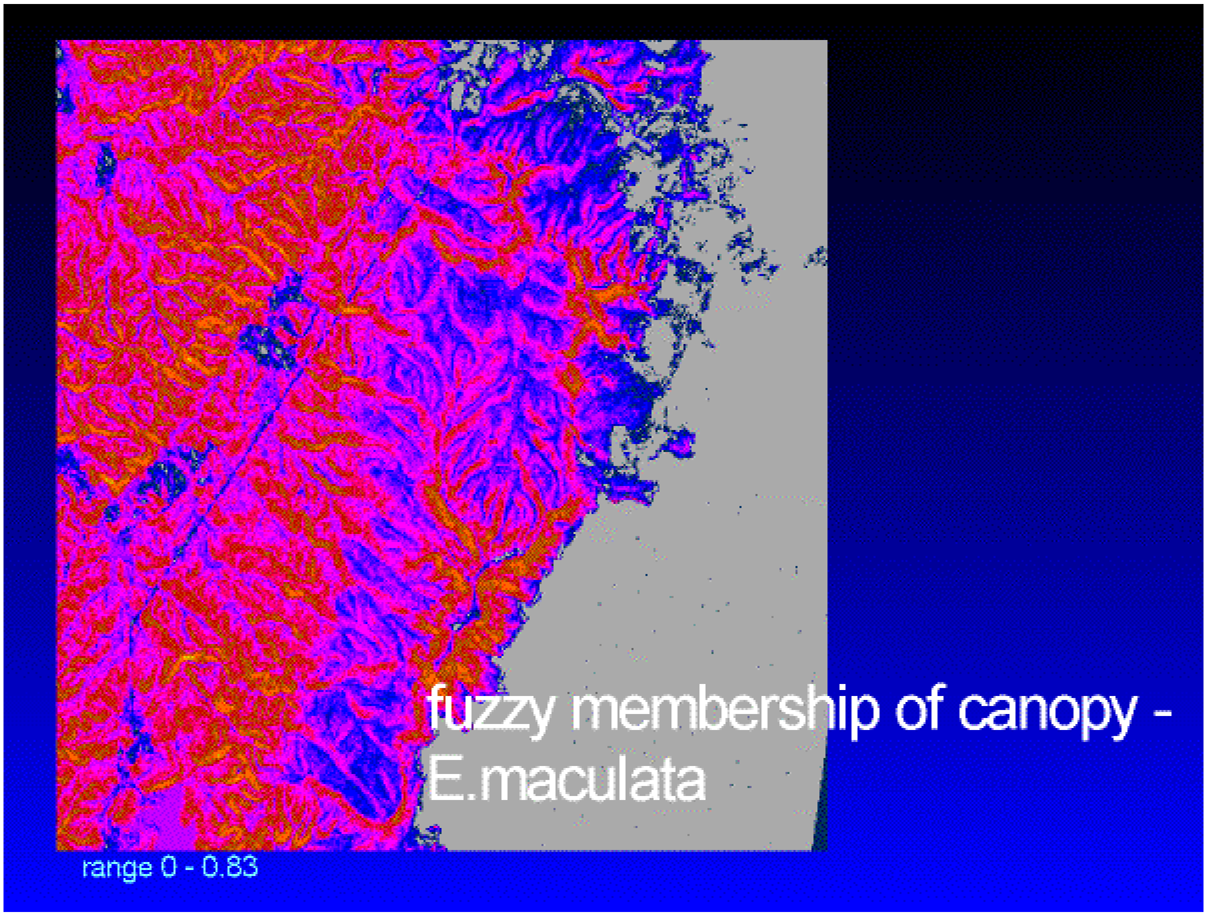


Figure 2: Fuzzy membership of the canopy for Eucalyptus maculata. Range is 0-0.83.

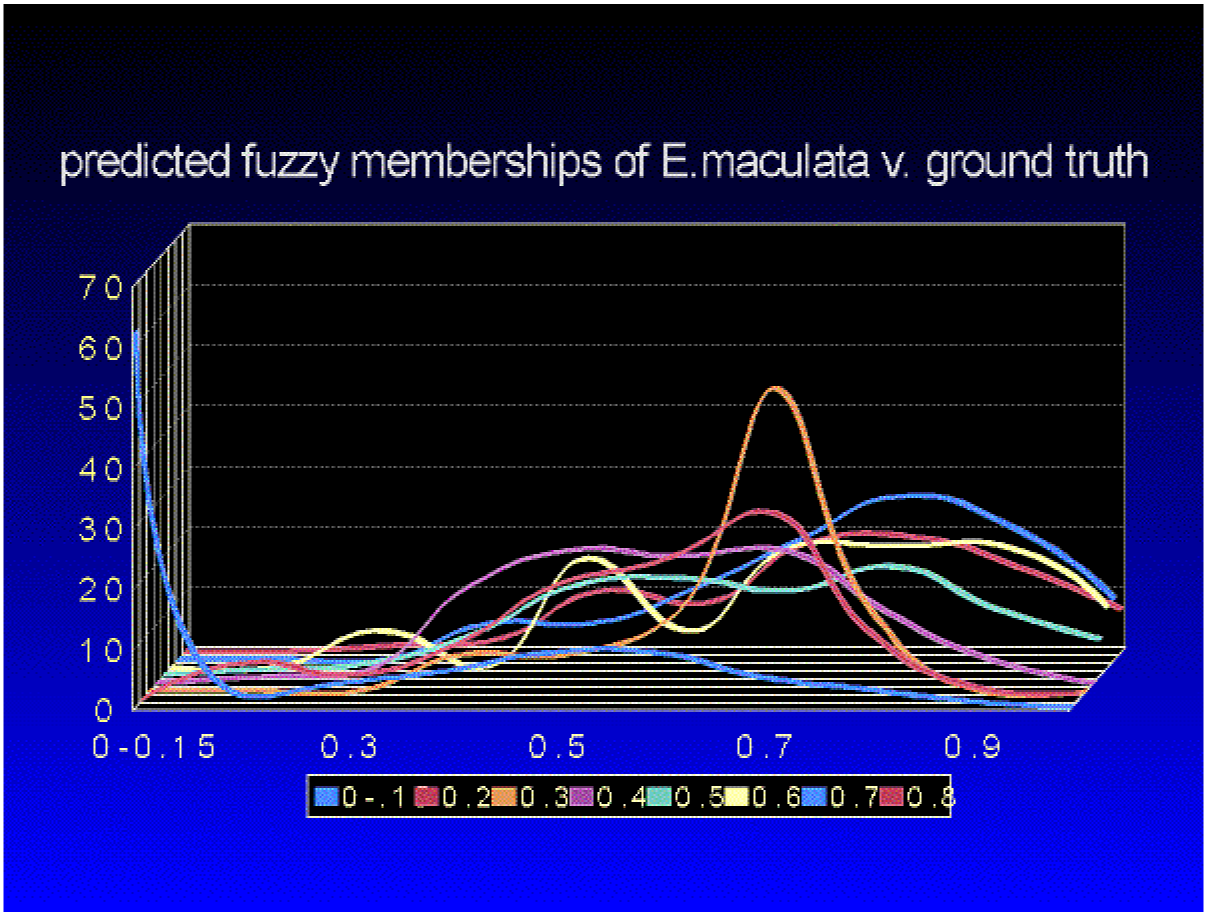


Figure 3: Distribution of predicted fuzzy memberships plotted against actual memberships. The broad pattern and structure are still being investigated.

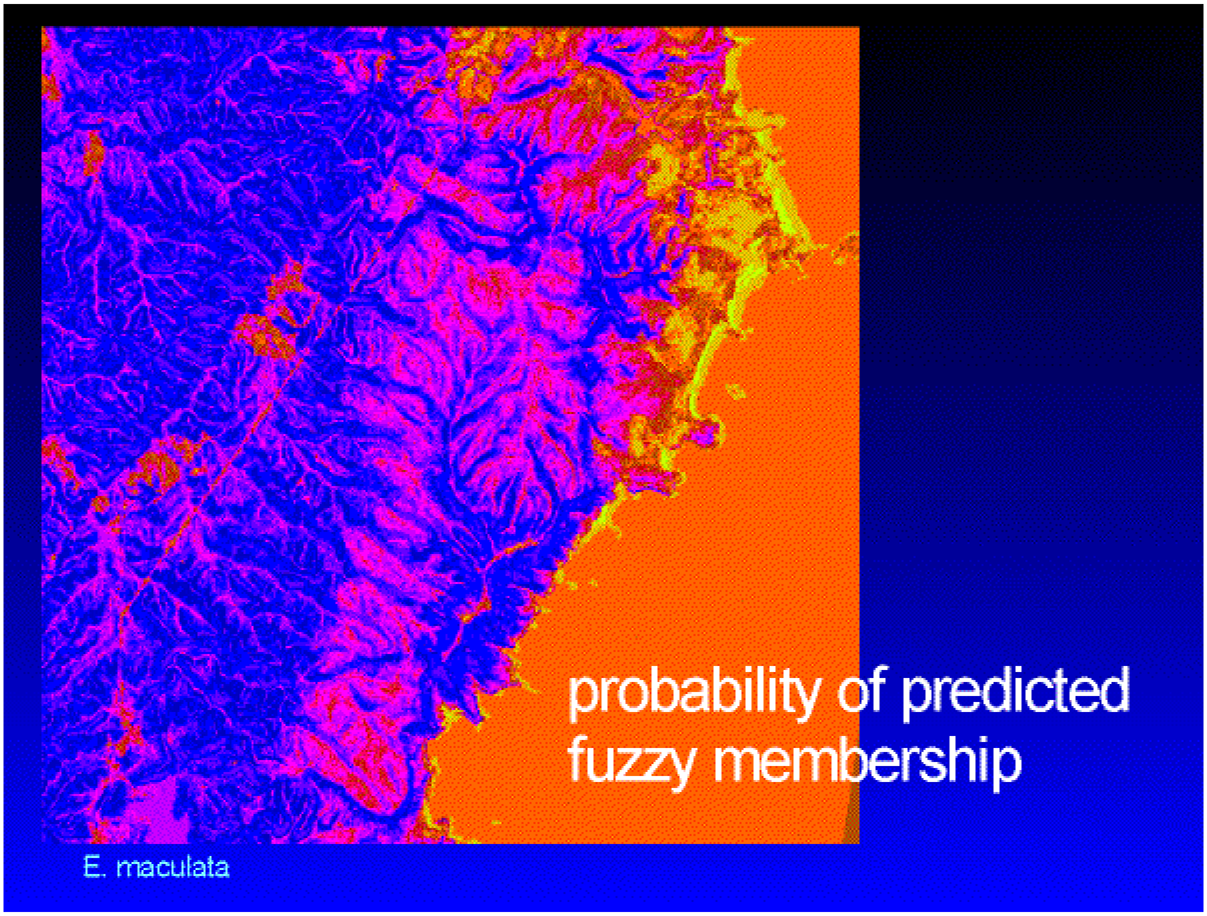


Figure 4: The probability of the predicted fuzzy memberships shown graphically. Absence has the highest probability and is the easiest to predict. The distribution of low values (blue) tend to be related to those areas of forest where gradients of change are flattest, and the high values (redder) where gradients are steepest.

The structure of the output is so different to our previous modelling exercises that comparative statistics are almost meaningless. The form of the model output is similar to that produced by Payne (Payne, Stockwell and Davey, 1994) using genetic algorithms on the Lees and Ritman (1991) dataset, but as that study only attempted to predict the probability of presence/absence for a species, it is not directly comparable. The information provided by this approach is much richer in content than a traditional choropleth map and would probably reside in a database rather than be made explicit as a map. For the species chosen as an example, and using only the fuzzy memberships, the RMS error is calculated to be 0.2324. Adding spatial and temporal context should reduce this significantly.

Conclusions

It is almost too simplistic to state that many of the compromises used by cartographers and geographers over the centuries need to be rethought in the light of modern technologies. The use of choropleth maps as the main data storage form of natural systems information has led to the perpetuation of an inappropriate data model for things such as vegetation, soil and geochemical mapping. It is easy to demonstrate that this is a major source of error but there is, as yet, no agreement on the most appropriate replacement. In this exercise we have examined one possible model which shows promise. Moving away from a `mapping' to a geographic information system mindset makes it clear how important it is to have an appropriate data structure, and how important it is to examine the data model to select the most appropriate data space for analysis.

References

Aspinall, R. & Lees, B.G. 1994. 'Sampling and analysis of spatial environmental data.' in Waugh, T. C. & Healey, R.G. (eds) Advances In GIS Research, Taylor and Francis, Southhampton, 1086-1099. ISBN 0-7484-0315-9 (B)

Austin, M.P. & Smith, T.M. 1989. `A new model for the continuum concept.' Vegetatio, 83: 35-47.

Fitzgerald, R.W. & Lees, B.G., 1993. `Assessing the classification accuracy of multisource remote sensing data.' REMOTE SENSING OF THE ENVIRONMENT, 47: 1-25.

James, P.E. 1967. `On the origin and persistence of error in Geography'. Annals of the Assoc. of American Geographers, 57:1-25.

Jupp, D.L.B., Walker, J. and Penridge, L.K., 1986. Interpretation of vegetation structure in Landsat MSS imagery: A case study in disturbed semi-arid Eucalypt woodlands. Part 2, Model-based analysis. J. Environmental Management, 23: 35-57.

Kauth, R.J. and Thomas, G.S., 1976. The Tasseled Cap. Proc. LARS 1976 Symp. on Machine Process. Remotely Sensed Data, Purdue University.

Lees, B.G. and Ritman, K. 1991. Decision tree and rule induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. Environmental Management, 15: 823-831.

Lees, B.G. 1994. `Decision trees, artificial Neural Networks and Genetic Algorithms for classification of remotely sensed and ancillary data.' Proceedings 7th Australasian Remote Sensing Conference, Remote Sensing and Photogrammetry Association, Australia Ltd, Floreat, W.A. v1: 51-60.

Lees, B.G., 1995. `Sampling strategies for machine learning using GIS' in GIS and Environmental Modelling: Progress and Research Issues, Goodchild, M.F., Steyart, L., Parks, B., Crane, M., Johnston, C., Maidment, D., and Glendinning, S. (eds), GIS World Inc, Fort Collins, Co. ISBN 1-882610-17-2. (E1).

Payne, K., Stockwell, D., and Davey, S. 1994. A methodology for improving the accuracy of vegetation mapping using GIS, remote sensing and genetic algorithms. Proc. of the regional conference of the International Union of Geographers: `Environment and the quality of life in Central Europe: problems of transition.' 22-26 August, Prague, Czech Republic.

Sinton, D.F. 1977. `The inherent structure of information as a constraint to analysis: mapped thematic data as a case study' International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems, Dedham, Mass. in Dutton, G. (ed), Harvard Papers on geographic information systems, Harvard University, Camb. Mass.

Brian G. Lees

Department of Geography

Australian National University

ACT 0200, Australia

e-mail Brian.Lees@anu.edu.au

Fax 06 249 3770

Telephone 06 249 3795

CONVERTING ADMINISTRATIVE DATA TO A CONTINUOUS FIELD ON A SPHERE

Waldo Tobler
Geography Department & NCGIA
University of California
Santa Barbara CA 93106-4060
tobler@geog.ucsb.edu

ABSTRACT

A procedure is developed for converting administrative unit (i.e., polygon) data, for example population, to a smooth continuous scalar field on a sphere. This reallocation from administrative units to a raster extends the pycnophylactic interpolation pro cedure previously developed for planar data to the spherical situation. It involves solving a partial differential equation in spherical coordinates by finite difference methods, and replaces the piecewise continuous discontinuities at the edges of the administrative polygons by a gentle transition, making the assumption that there is autocorrelation in the data. An example is presented.

INTRODUCTION

A recent project (Tobler, et al 1995) required the reallocation of sub-national populations to a grid of latitude and longitude quadrilaterals. This reassignment of the populations from the political or census units to a set of spherical cells provides a useful format for many modeling tasks. The data are then also more easily matched to geophysical and earth satellite observations. Conversion from administrative units to ecologic zones or drainage basins is similar.

In this procedure country or census boundaries are first recorded as polygons defined by latitude & longitude coordinates. The population figures for these units are then assembled. Next a lattice of uniform cells is positioned over the political units, invoking a polygon to grid procedure. The population is then evenly spread over the cells of each unit. This yields what can be thought of as a piecewise continuous surface of population, with discontinuous jumps at the political unit boundaries. To provide more realism a redistribution within each unit is now undertaken. But this reallocation must preserve the original values within each region. Thus the primary condition for mass preservation is the invertibility condition for any method of areal information redistribution:

$$\iint_{R_i} f(x, y)\, dx\, dy = V_i \quad \text{for all } i,$$

where Vi denotes the value (population in the present context) in region Ri (a subnational polygon). In the spherical case the integral is modified to take into account the convergence of the meridians. Many functions can be found which satisfy this condition so that additional constraints must be supplied. The one chosen is designed to give an approximation to a smooth continuous surface of proper content. It considers the populations of adjacent units and the density transition between units becomes sm ooth and continuous. This type of data redistribution was introduced for a planar set of regions in a previous publication (Tobler 1979a) and exists in a computer program available from the AAG Microcomputer Specialty Group. The procedure is sometimes known as areal interpolation since the data exist only for areas and are not given at point locations. The present note extends this procedure to a spherical domain. Other geographic applications in which spherical versions of similar equations appear are br iefly cited.

THE PROCEDURE

The reallocation is required to result in a smooth discretized surface whose volummetric content is identical to the initially given value when integrated over the comparable irregularly shaped region. In the discrete case the integration is of course a summation over the nodes of the grid or lattice. The steps in the procedure are to 1) for each of the separate polygons assign and distribute the entire regional population to the cells of a grid covering that polygon; then 2) reallocate the population smo othly to these same cells in each region, taking into account the population values in the neighboring regions, and then 3) test whether the population totals within each region are still correct, and if not, to apply a correction factor. The smoothing uses a converging iterative algorithm that requires many passes over the lattice. The only part of this procedure that requires modification for the spherical case is the smoothing operation in step two, and this is what is covered here.

If one were to apply the procedure to all of South America, using the populations of the countries (defined by polygons bounded by latitude and longitude points) as the data to be redistributed, then it might be thought that an adjustment would be needed for the first step in the procedure. Chile, for example, is a long north-south country and the spherical quadrilaterals covering the country are of unequal area, shrinking in size toward the South pole. The assignment of the population to the cells could then be adjusted for this latitudinal variation of the latitude/longitude quadrilateral sizes. This does no harm but is not really necessary. The iterative procedure used for the spherical smoothing adjustment overrides the initial pattern. All of the pop ulation could actually be assigned to just one cell in the lattice within each country and the final result would be the same, although the iterations might take a bit longer. An exception should be noted here. If the objective is simply to spread the dat a evenly over the cells of a region representing a polygon, without any subsequent smoothing, then the latitudinal variation in the cell sizes needs to be taken into account. Otherwise the data are not spread evenly, increasing in density as the cells be come smaller. To simplify things this adjustment is recommended in all cases. The smoothing step itself does need modification to incorporate the changing spherical cell sizes.

A smooth function, intuitively, is one that has few local oscillations. Stated in another way, neighboring locations should have similar values. Such autocorrelation seems to exist for many human distributions. This is an assumption about spatial demographic processes, a form of geographic insight capturing the notion that most people are gregarious and congregate. Densities, and other characteristics, in neighboring areas therefore tend to resemble each other unless there is a physical barrier. Thus one can say that there is only a small rate of change in all directions. Mathematically this requires that the functions' partial derivatives be small. Consequently it seems natural to minimize the sum of the squares of the partial derivatives, i.e., minimize:

$$\int\int \left(\frac{\partial U^2}{\partial x} + \frac{\partial U^2}{\partial y} \right) dx\, dy \;.$$

Here the integral is to be taken over the entire region of interest, and U(x,y) is the function in question. The solution to this least squares problem is another equation, known as the Laplace equation (Kantorovich and Krylov 1958, p 246 et seq.):

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0 \;.$$

Solving this new partial differential equation, with appropriate boundary conditions, gives the required smooth function U(x,y). To solve the equation on a digital computer it is usual to discretize the equation and to work with a finite lattice. The fini te difference version to Laplace's equation is obtained by using the well known approximations

$$\frac{\partial^2 U}{\partial x^2} \approx U_{i,j-1} - 2U_{i,j} + U_{i,j+1} \;, \quad \frac{\partial^2 U}{\partial y^2} \approx U_{i-1,j} - 2U_{i,j} + U_{i+1,j} \;.$$

Here we use i for a row subscript and j for a column subscript with the value at each node of the mesh labeled U and subscripted by its row and column position. Adding these two equations and setting the result equal to zero, yields the condition that the value at any mesh point must be equal to the average of its neighbors. This is another way of defining what is known technically as a 'harmonic' function. Expressed in symbols, using four grid neighbors, this gives the equation:

$$U_{i,j} = ¼\,(U_{i-1,j} + U_{i+1,j} + U_{i,j-1} + U_{i,j+1})$$

An even stronger requirement is that some derivative be the average of its neighbors. This results in a higher order of smoothness. The usual approach is to use the Laplacian of Laplace's equation. After some tedious but simple algebra and taking the grid spacing Dx = Dy = 1, the required condition becomes

$$U_{i,j} = [\, 8\,(U_{i+1,j} + U_{i-1,j} + U_{i,j+1} + U_{i,j-1}) - 2\,(U_{i+1,j+1} + U_{i+1,j-1} + U_{i-1,j+1} + U_{i-1,j-1}) - (U_{i+2,j} + U_{i-2,j} + U_{i,j+2} + U_{i,j-2}) \,] / 20.$$

This is a finite difference approximation to the biharmonic equation

$$\frac{\partial^4 U}{\partial x^4} + 2\,\frac{\partial^4 U}{\partial x^2 \partial y^2} + \frac{\partial^4 U}{\partial y^4} = 0$$

which solves a least squares problem minimizing the surface curvature. These are the two alternate smoothing operations proposed and previously described for the mass preserving interpolation problem on a plane. Boundary conditions are also needed for a solution.

SPHERICAL VERSIONS

The Laplace equation on the surface of a unit sphere is

$$\frac{1}{\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial U_{\phi,\lambda}}{\partial\phi}\right) + \frac{1}{\cos^2\phi}\frac{\partial^2 U_{\phi,\lambda}}{\partial\lambda^2} = 0$$

$$\phi = latitude \quad , \quad \lambda = longitude$$

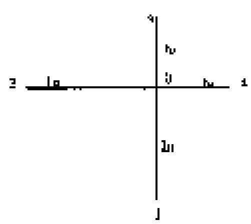
and the spherical surface biharmonic equation is

$$\frac{1}{\cos^2\phi}\frac{\partial}{\partial\phi}\cos\phi\frac{\partial^2}{\partial\phi^2}\cos\phi\frac{\partial U}{\partial\phi} + \frac{1}{\cos^3\phi}\left(\frac{\partial^3}{\partial\phi\partial^2\lambda}\cos\phi\frac{\partial U}{\partial\phi} + \frac{\partial}{\partial\phi}\cos\phi\frac{\partial^3 U}{\partial\phi\partial^2\lambda}\right) + \frac{1}{\cos^4\phi}\frac{\partial^4 U}{\partial\lambda^4}$$

To get the spherical versions in discrete form, begin with the harmonic property and modify the finite difference approximation to the Laplace equation so that it can be applied on the surface of a sphere. That is, use the fact that the value at any location is the average of its neighbors. For this purpose a weighted average of the neighbors must be used because the neighbors are at different spherical distances in the latitude/longitude graticule.

The general centered difference formula for the Laplace equation when the central point at u0 is related to neighbors at different distances is given by Rektorys (1969, p. 1114), and is explained by reference to the diagram:

$$\frac{2}{h_1+h_3}\left(\frac{u_1-u_0}{h_1}+\frac{u_3-u_0}{h_3}\right) + \frac{2}{h_2+h_4}\left(\frac{u_2-u_0}{h_2}+\frac{u_4-u_0}{h_4}\right) = 0 \, ,$$



For the spherical case the distances h1 and h3 along the meridian are equal to each other, and h2 and h4 (along a parallel) are equal to each other but these are not equal to the meridional distance. Let h1 = h3 = h and h2 = h4 = h cos(f), from the geometry of the sphere. Then the sought after average value at Ui,j is obtained from some simple algebra and is, reverting to the row, column notation:

$$U_{i,j} = \frac{1}{2}\left[U_{i,j-1} + U_{i,j+1} + \cos^2\phi\,(U_{i-1,j} + U_{i+1,j})\right] / (1 + \cos^2\phi) \, .$$

The divisor is a normalization factor to take into account the unequal weights in the weighted average. This agrees with the result by Swarztrauber (1974), and approximates that of Gates & Riegel (1963), if we consider that the values apply to small quadrilaterals not near either pole. It assumes that the cosine values vary by only a small amount over the small increment. The cosine value should be taken at the center of the cell and not at the edge. These equations further assume a sphere of unit radius and equal increments in latitude and longitude. If the spacing is not equal or the cells are of large size then the more general equations must be substituted. The biharmonic case takes a bit more algebra but uses the same logic. A similar, slightly more complicated, modification is required to adapt the equations to an ellipsoid.

In the computer program one needs to know the latitude of the southernmost (or northernmost) mesh line and the latitudinal increment between these lines. A table of squared cosines and divisors is precomputed so that these do not need to be recalculated on every iteration. If the cell centers are to contain the information then these cosine values should be shifted by one half the increment. It is also necessary to know whether the domain extends over 360 degrees of longitude. If it does then the computat ion must recognize the periodic nature of the east-west boundary. Otherwise a Dirichlet or Neumann condition must be specified at all edges. In the context of the current application the grid will not reach the poles since there will be no data of relevance there.

FURTHER APPLICATIONS

The smooth pycnophylactic reallocation for areal data can also be used as one step in the more general conversion of values from one set of polygons to another, e.g., converting information from political units to ecological zones, or vice versa. The modification described here allows this to be done when the two sets of polygons are defined by latitude/longitude coordinates. All that is needed is a reaggragation over all of the cells of the new set of polygons. The magnitude of any error depends on the r esolution given by the cell size.

Several additional uses of these equations can be noted. In Tobler & Kennedy (1985) a procedure was described in which the Laplace equation (or the biharmonic equation) is used to solve the interpolation problem for randomly scattered points on a plane. Results given here allow this interpolation technique to be applied to data irregularly arranged on a sphere. Also see Tobler (1979b and 1994) for related uses of smooth interpolation from values given at point locations. In Dorigo and Tobler (1983) an alg ebraic model of geographic movement is developed which allows estimation of a potential function and associated vector field from empirical migration data by solving Poisson's equation. The finite difference approximation to the Poisson equation is identi cal to that used for the Laplace equation aside from the term defining the source/sink field. Using f(f,l) as this forcing function one obtains

$$U_{i,j} = [U_{i,j-1} + U_{i,j+1} + \cos^2(\phi)\,(U_{i+1,j} + U_{i-1,j}) + \cos^2(\phi)\, f(\phi,\lambda)\,] \;/\; [2(1 + \cos^2(\phi))].$$

The push-pull model can therefore be applied to movement data on the surface of the earth, assumed spherical, with observations in a small mesh of uniform latitude/longitude increments.

EXAMPLE

The example given here uses Tasmania, since it is relatively small, and only eight polygons are involved. The data used are the estimated 1994 population counts for these eight administrative units. The tables give the population before and after the reallocation towards smoothness. The adjustment for uniform density on the sphere has only a minor effect since

Tasmania extends over only circa 4 degeees of latitude. The reallocation taking into account values in each regions' neighhbors is however quite d ramatic. Further examples are shown in Tobler, et al 1995.

References:

AAG Microcomputer Specialty Group, 1994, Spatial Sciences Center, Indiana University of Pennsylvania, 212 Whitmyre Hall, Indiana, PA 15705-1087.

Dorigo, G., and W. Tobler, 1983, "Push-Pull Migration Laws", Annals, Assoc. Am. Geographers, 73,1:1-17.

Gates, W., & C. Riegel, 1963, "Comparative numerical integration of simple atmospheric models on a spherical grid", Tellus, XV:410.

Kantorovich, L., and V. Krylov, 1958, Approximate Methods of Higher Analysis, Nordhoff Interscience, The Hague.

Rektorys, K, ed., 1969, Survey of Applicable Mathematics, MIT Press, Cambridge.

Swarztrauber, P., 1974, "The Direct Solution of the Discrete Poisson Equation on the Surface of A Sphere", J. Computational Physics, 15: 46-54.

Tobler, W., 1979a, "Smooth Pycnophylatic Interpolation for Geographic Regions", J. Am. Stat. Assn., 74, 367:519-536.

Tobler, W., 1979b, "Lattice Tuning", Geographic Analysis, 11,1:36-44

Tobler, W., and S. Kennedy, 1985, "Smooth Multidimensional Interpolation", Geographical Analysis, 17,3:251-257.

Tobler, W., 1994, "Bidimensional Regression", Geographical Analysis, 26:186-212.

Tobler, W., U. Deichmann, J. Gottsegen, K. Maloy, 1995, The Global Demography Project. Technical Report TR-95-6, NCGIA, Geography Department, University of California, Santa Barbara, 75pp.

Tables

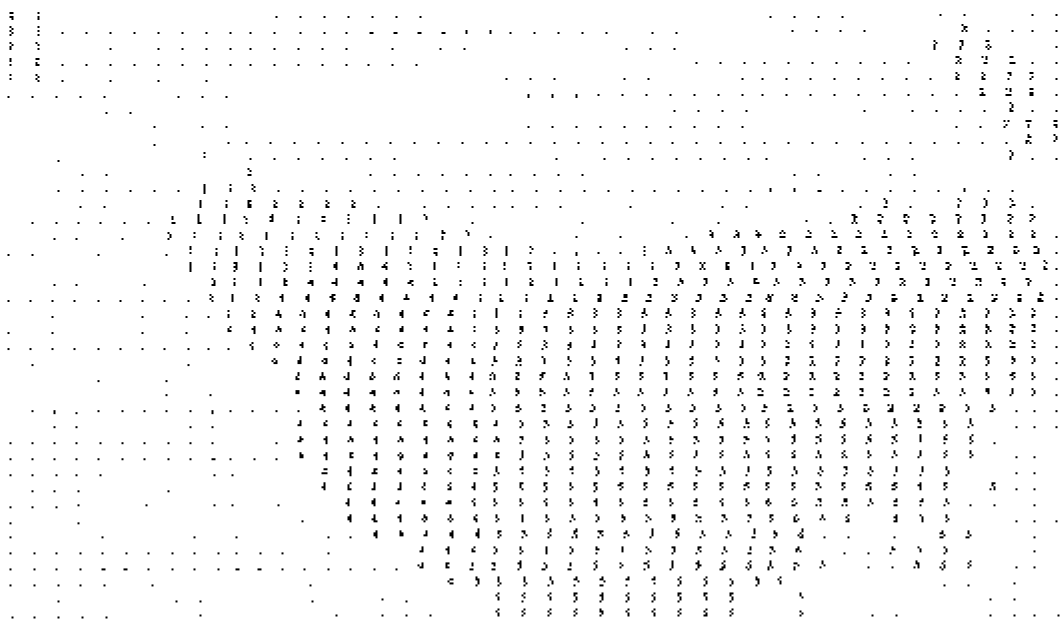
Tasmania: The tables are 40 rows by 45 columns in size and contain population estimates for 0.1 degree quadrilaterals obtained from the eight regions listed below. The SW corner of the array is at Lat -43.6, Lon 143.9. The grid size is 1/10 of a degree (6 minutes) in both latitude and longitude. The first table gives the zone number, one to eight, for each quadrilateral. Outside of Tasmania no data are given. The grid was created by a polygon to raster program using polygons describing the Tasmanian regions using latitude and longitude coordinates for the polygon vertices.

The Tasmanian data refer to the following zone estimates:

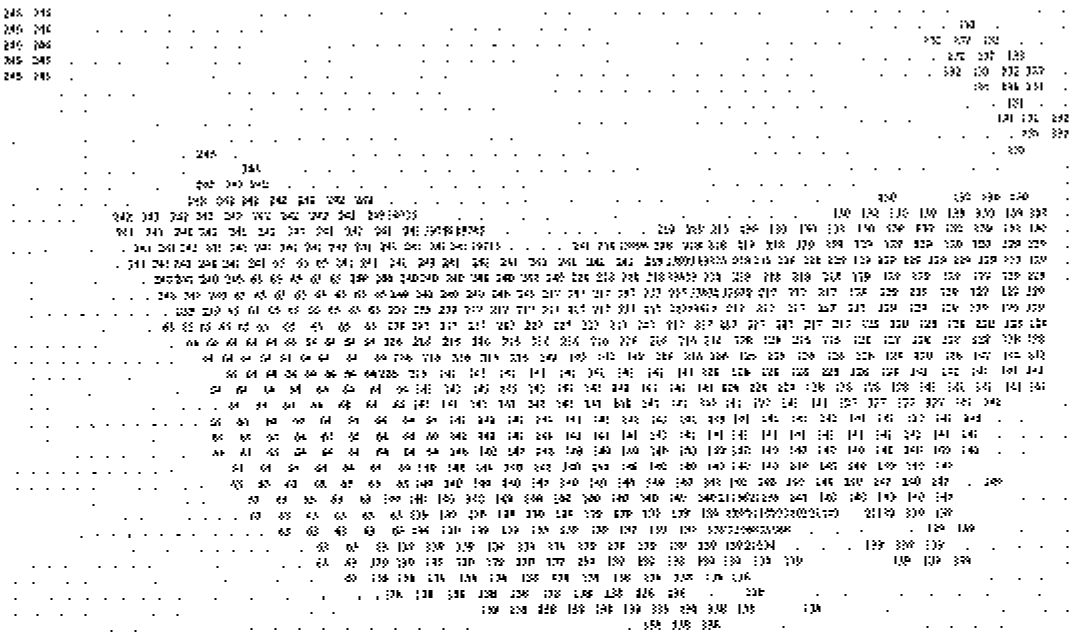
1. North Western rural, 24859 people, 9593 sq. kilometers
2. North Eastern, 16275 people, 11994 sq. kilometers
3. Central North 186 71 people, 7900 sq. kilometers
4. Western, 8898 population, 12746 sq. kilometers
5. Southern, 36918 population, 23820 sq. kilometers
6. Greater Hobart, 189850 persons, 937 sq. kilometers
7. Burnie-Devonport, 78979 people, 579 sq. kilometers
8. Greater Launceston, 97672 persons, 801 sq. kilometers

Totals: 472,122 persons, 68,370 sq. kilometers

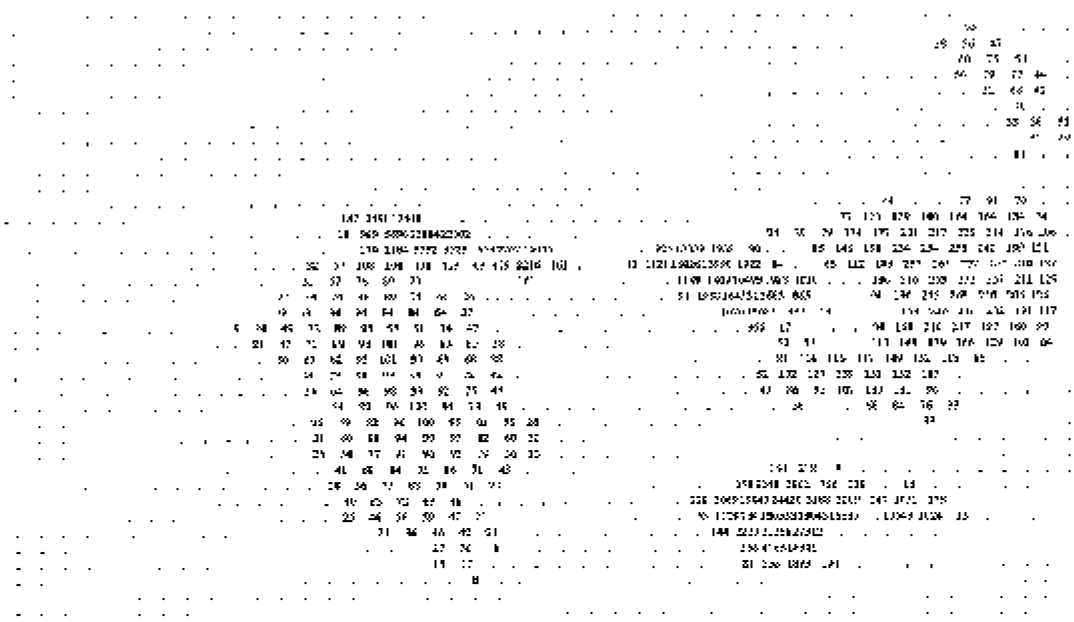
Tasmania: Identifying Array - 8 regions



Tasmania: uniform density - taking into account the spherical shape. The numbers are the number of persons in each 0.1 degree quadrilateral.



Tasmania: smoothed density - taking into account the spherical shape. The numbers are the number of persons assigned to each 0.1 degree quadrilateral.



Steven P. Frysinger

Dynamic Finite Difference Grid Generation for Environmental Decision Support Systems

The information-intensive nature of environmental management has led to increased interest in a class of computer applications called Environmental Decision Support Systems (EDSS). These are systems designed to assist humans making complex environmental management decisions, employing multiple technologies for this purpose. An open architecture for an EDSS family integrating mathematical models, Monte Carlo simulation, and GIS was previously reported (Frysinger et al 1993a, Frysinger et al 1993b, Frysinger 1995). One element missing from this architecture was a mechanism for supplying the finite difference models (commonly used in EDSS) with reference grids and boundary conditions properly oriented in the x-y plane and/or populated with boundary values when conditions differ from one simulation iteration to the next. To account for variations in grid orientation or boundary value arrangement from iteration to iteration, one might employ a dynamic method for generating and populating a new grid for each iteration of the simulation. The present work addresses such a technique, in which one first computes grid positions, overlaying grid cells onto the EDSS' GIS data, and then computes boundary values for each cell by reference to these data.

Introduction

An architecture for Environmental Decision Support Systems (EDSS) integrating mathematical models with Geographic Information Systems (GIS) was previously described (Frysinger et al 1993a, Frysinger et al 1993b, Frysinger 1995). This architecture employs Monte Carlo methods to simulate uncertain parameters in the conceptual model under evaluation, and within each iteration of the Monte Carlo loop one or more mathematical models of environmental processes are solved.

For example, when simulating the leak of materials from a low-level radioactive waste facility, each iteration might include a separate, serial execution of models describing the nature of the release, migration through the unsaturated zone, transport in the saturated zone, passage into a surface water body, uptake by consumers of the surface water, and allocation to the various exposed organs of the simulated individual.

This is an open software architecture, referring to a software framework which can readily be modified or extended. Beyond this, the architecture emphasizes the use of public domain components to help make the resulting software accessible to those who need it, resting on a Unix workstation platform under the X Window System, and using the GRASS GIS.

GRASS's analytical tools are invoked through Unix's system call or its equivalents, with data from files moved between the GIS domain and the modeling codes via shell scripts under a data passing protocol. Spatial animation data (such as contaminant plumes or hypothetical wells) are drawn onto the GRASS window without the intervention of GRASS's native code, yielding a much more interactive spatial display than is provided by the conventional GIS.

Individual modeling codes are maintained as stand-alone programs managed by shell scripts. Data import and export for these codes are accomplished by providing data to the Unix standard input file, and trapping results from the standard output file. Alternatively, codes which use named files for input and output are provided with names of temporary files for this purpose. These data can then be reformatted using standard Unix tools (such as awk and sed) to conform to the requirements of subsequent codes. When Monte Carlo simulation techniques are to be used to manage parameter uncertainty, the simulation loop itself is implemented within a shell script simulation manager, which also manages the interconnections and data transformations necessary to provide data flow between modeling codes.

This arm's length relationship with computational codes allows existing codes, such as those from various Federal agencies, to be used "as is", with no modifications. Apart from cost reduction, this preserves the status of codes which have already been submitted to a quality assurance process. This desire to leave "off-the-shelf" codes unchanged is behind the present work.

In some cases, environmental processes are modeled by finite difference methods, wherein systems of equations are solved within each cell of a grid constructed for the purpose. In such a case, each iteration of the Monte Carlo simulation involves a complete solution of the finite difference model on the grid, which has been initialized with boundary conditions and initial values appropriate to the model. Computer codes implementing such finite difference models generally require the user to provide the geometry of the grid and the initial values as inputs to the program. If the same grid and values can be used for all iterations of the Monte Carlo simulation, then its development can be treated as part of the conceptual model development process, generally accomplished with a fair amount of human intervention.

Difficulty arises, however, when one of the uncertain parameters being varied from iteration to iteration of the Monte Carlo process changes the preferred orientation and/or initial contents of the finite difference grid. For example, the path or width of a simulated stream may vary from iteration to iteration (as a result of other uncertain parameters), influencing the orientation and/or cell population of a grid which must include constant head values representing the interface between the stream and the groundwater. Since the grid configuration must change based upon pseudorandom sampling, the grid can't be statically populated with boundary conditions and initial values. And since such Monte Carlo simulations often involve hundreds or thousands of iterations, it is impractical to consider manual development of all possible grid configurations.

But if the grid values can be mathematically derived from spatial data represented in the underlying GIS, it is possible to dynamically generate the grid and populate its cells by reference to the appropriate GIS layers. The simple method proposed here will assume the use of a raster GIS (such as GRASS) and a uniform finite difference grid configuration.

Method

In the most general case, it will be necessary to rotate the finite difference grid between simulation iterations, as well as to adjust the cell dimensions (which are here assumed to be uniform, but which could with additional attention vary over the simulated field). This transformed grid must then be overlaid onto one or more GIS layers (which themselves might have been computed dynamically and thus differ with each iteration). The proposed method therefore involves two steps:

1. Transform (i.e. rotate and scale) the grid to correspond to the desired new orientation and pattern
2. Populate each cell in the transformed grid by computing the initial value based upon the appropriate mathematical relationship operating over the aggregate of GIS values contained within the confines of that cell

The first step can be accomplished using conventional two-dimensional coordinate transformation procedures (such as is commonly employed in computer graphics).

Let grid G(i,j) be defined to have its origin, G(0,0) at GIS coordinates (x0,y0). Assume uniform cell dimensions of (Di,Dj), and let A represent the angular difference between the grid axis and the underlying GIS axis.

Define operator L[G(i,j)] as the GIS coordinate of grid intersection (i,j), such that L[G(0,0)] = (x0,y0). Thus, Lx[G(i,j)] = x0 + iDicos(A) - jDjsin(A), and Ly[G(i,j)] = y0 + iDisin(A) + jDjcos(A). This operator can now be used to compute the GIS coordinates of the corner points describing the cell (i,j), namely G(i-1,j-1), G(i,j-1), G(i-1,j), and G(i,j).

If the grid resolution is much finer than the GIS layer resolution, it is possible (but not necessary) that all four cell corners lie within the same GIS raster cell. In this case, it is a simple matter to compute the grid cell's initial value by reference to this single GIS cell (assuming the corresponding relationship has been defined). For example, returning to our shifting stream example, if the grid cell falls entirely within a GIS cell or polygon which is within the stream, then the grid cell can be given be given the fixed head value corresponding to the stream surface altitude (which is in turn a combination of streambed altitude and stream depth).

However, in the more general case, the grid cell will intercept or include more than one GIS layer cell, and may (if the grid cell resolution is much less fine than that of the GIS layer) intercept a large number of GIS cells. In this case, one must compute a grid cell initial value based upon an aggregation of the included or intercepted GIS cells.

The simplest way to accomplish this is by using the tools of the underlying GIS. Considering the grid cell as a polygon (defined by its four corners, whose GIS coordinates were computed above), most GISs provide a means of reporting the average GIS layer value within the polygon (e.g. by converting the cell's polygon to a raster mask). This average value can then be used in the calculation of the grid cell's initial value. So, returning for the last time to our stream example, if the grid cell contains GIS values whose average is biased towards "water", then we would assign that cell the constant head value described earlier.

Conclusion

A simple method has been proposed with which finite difference grid cells can be defined and initialized dynamically with reference to GIS data. This technique will be especially useful in Environmental Decision Support Systems which integrate finite difference models into Monte Carlo simulation loops. In such systems there are several reasons why a single finite difference grid cannot be predefined for all iterations of the loop, including changing flow orientations and changing base data. In such situations, dynamic grid generation can be performed for each iteration of the simulation, and data from the underlying GIS (which may themselves have changed between iterations) can be used to populate the grid with boundary conditions and initial values.

References

Frysinger, S.P., Thomas, R.P, and Parsons, A.M. (1993a) Hydrological Modeling and GIS: the Sandia Environmental Decision Support System. In K. Kovar and H.P. Nachtnebel (editors), *Applications of Geographic Information Systems in Hydrology and Water Resources Management - Proceedings of HydroGIS 93*, IAHS Press, April 1993.

Frysinger, S.P., Copperman, D.A., and Levantino, J.P. (1993b) Environmental Decision Support Systems: An Open Architecture Integrating Modeling and GIS. In *Proceedings of the Second International Conference on Integrating Geographic Information Systems and Environmental Modeling*, NCGIA, September 1993.

Frysinger, S.P. (1995) An Open Architecture for Environmental Decision Support. *International Journal of Microcomputers in Civil Engineering, Vol. 10, No. 2*.

Steven P. Frysinger, Ph.D.
Associate Professor
Director, Center for Applied Water Technology
James Madison University
College of Integrated Science and Technology
Harrisonburg, Virginia 22807
Tel: 540/568-2710
Fax: 540/568-2761
Net: frysinsp@jmu.edu

Catherine Gautier and Pete Peterson

Issues Linked to Geographical Information Systems in Global Environmental Research: Data Base Handling and Multi-Sensor Data Fusion

---

1. Introduction

With the enormous flow of data that will be generated by the sensors on the upcoming US and international Earth Observing platforms, a number of data base-related issues will have to be tackled if we want these new data sets to be optimally used for global studies. While some of those issues are being addressed at the program level such as with EOSDIS (see other papers at this conference), other issues need to be looked at more from a users' perspective. This is what is done here. Among those data base issues for the users to handle is the fusion of data coming from different sources: satellites and orbiting platforms and in-situ (ship, aircraft, surface measurements), and their integration into models. One way to approach this issue is 4-D data assimilation whereby general circulation models of the atmosphere and the ocean are used to "assimilate" the data. This is however not a unique approach and may not be the best one to unravel new processes, because these processes can be masked by all the other processes that are at work within the complex model used in the assimilation.

The approach we discuss here is one by which enormous data sets generated from different sources are combined and integrated with simpler models to produce low level geophysical parameters for which we do not have an existing data base at a high time and space resolution. This is generally true for geophysical processes dealing with the ocean, or the atmosphere over the ocean, where the sampling in both time and space is too limited. We have however recently understood the crucial role these oceanic and air-sea interaction processes play in the variations and evolution of the earth's climate and therefore, the need to improve our knowledge of their variability.

The main questions we will address here are: (1) how do we extract scientifically useful information from 10's of Gigabytes of satellite radiance data and first order geophysical products which have variable time and space sampling? and, (2) what role can GIS tools play in this data melt-down whereby hundreds of 2-D data sets over long periods of time are essentially digested down to a single curve ?

To focus our approach to these questions, we discuss an example of a scientific topic that can presently be addressed with available data and faces many of the same issues as those that we will be faced when EOS and other enormous data sets become available, at the end of the century. The scientific question we are interested in investigating is that of the annual mean meridional oceanic heat transport. This scientific topic is of extreme importance to climate since the oceans play a major role in redistributing the heat that is acquired at low latitudes and lost at high latitudes. This heat redistribution effected by the major ocean currents helps drive weather and climate on the adjacent continents due to the exchanges of heat that take place between the ocean and the atmosphere at all latitudes.

We first describe the data manipulation we presently perform to derive the annual mean meridional oceanic heat transport, including its various components Then we discuss the areas within which GIS tools could facilitate our computations. Finally we investigate the areas for which GIS tools could have a unique contribution.

2. Meridional Oceanic Heat Transport Computations

There are many ways to compute the meridional oceanic heat transport, the most direct one being to use ocean currents together with ocean temperature distribution. This however requires knowledge of the vertical distribution of currents and temperature over all the oceans at a spatial resolution commensurate with currents and temperature spatial variability and a temporal resolution of at least a season to obtain the mean annual meridional heat transport. Data at these resolutions have not yet been available and therefore our knowledge of the meridional oceanic heat transport is limited to long-term (10-50 year) averages. Many indirect approaches are possible and here we present one whereby the net surface heat flux is first computed over all the oceans and the meridional oceanic heat transport is estimated from the meridional divergence of the zonal mean net surface heat flux. The main issue is therefore to compute the net heat flux at the ocean-atmosphere interface. This flux, Qnet, is given by:

$$\text{Qnet} = \text{SW} - \text{LW} - \text{LHF} - \text{SE}$$

Where SW is the net surface solar (shortwave hence, SW) radiation flux, LW is the net longwave radiation flux, LHF is the turbulent latent heat flux corresponding to the evaporation of water from the ocean surface and SE is the turbulent sensible heat flux. The net heat flux is taken to be positive going into the ocean and negative when leaving the ocean. Each term is evaluated individually but the two largest terms, SW and LHF will be focused on here. The other two terms are temporarily obtained from climatologies.

2.1 Latent Heat Flux (LHF)

The latent heat flux is a high level geophysical parameter which is directly measurable neither from the surface nor from space. It is typically estimated using what is called a parameterization, which is a formulation that relates a parameter at a generally small scale to some "measureables" at a larger scale. In the case of the latent heat flux, the "measureables" are the low level geophysical products: surface wind speed, relative and absolute humidity at the ocean's surface and surface ocean temperature. These can be measured from ships, but are not directly measurable from space. They can, however, be estimated from microwave and infrared radiance data through the application of a retrieval algorithm.

The determination of latent heat flux over the ocean surface from space is based on microwave brightness temperatures from the SSM/I sensor which flies on the DSMP satellites series and infrared radiances from the AVHRR sensor on the NOAA satellites series (Jourdan and Gautier, 1995). The microwave brightness temperatures are transformed into surface wind speed and total precipitable water in the atmosphere, among other geophysical parameters by F. Wentz after a number of data manipulations (Wentz, 1983, 1992). The AVHRR infrared radiances are transformed into sea surface temperature by NOAA (Smith, 1992). All these data manipulations (application of operational retrieval algorithms) are similar to those that will be applied to EOS data in the near future and are not discussed here. We only start the discussion of our analysis with the low level geophysical data sets which are available to the scientific community.

2.2 Shortwave Radiation (SW)

The SW product is, like the LHF, based on a number of satellite data sets. In this case the data base of origin is a set of satellite radiances in the visible and infrared spectral regions taken from geostationary and polar orbiting satellites. The data are remapped over the globe (Rossow, 1991) and analyzed to produce cloudy and clear (surface) radiances as well as percentage area covered by clouds. Together with sun-illumination and viewing geometry and atmospheric properties (such as total precipitable water and total ozone amount), these data can be used to compute the net solar surface radiation (or shortwave radiation). Daily and monthly fields of SW can be computed (Gautier et al., 1980) and included in the overall computations of the net surface heat flux.

2.3 Meridional Oceanic Heat Transport

LHF and SW are combined with climatological values of LW and SE to produce monthly mean fields of surface heat flux. The meridionial oceanic heat transport is computed by integrating from north to south the zonal average of the net heat flux. A simple rescaling by the cosine of the latitude is necessary to obtain the actual flux values.

3. Data Base Manipulation and Multi-Sensor Data Fusion

The selected application - annual mean meridional oceanic heat transport - can easily be generalized to many climatologically relevant products. The procedure used applies a number of data manipulation steps to derive a high level geophysical climatological product from a low level geophysical products that have been obtained from multiple satellite sensor or in-situ measurements. The high level geophysical product thus obtained have two characteristics: (1) they are many steps removed from the direct satellite data or in-situ measurements and, (2) they represent a new type of parameters that has never been derived before at this resolution. Its characteristics and time and space variability, therefore, are little known. These imply that many quality and accuracy evaluations must take place at every step of the computation process. While the quality evaluation is generally performed through visualization, the accuracy evaluation is usually performed through some type of statistical analysis.

3.1 Data Base Manipulation

Even though the data volume has already been reduced by at least an order of magnitude when we acquire it, because we are interested in climate problems, the data sets need to extend over long periods of time. We, therefore, deal with very large volumes of data. In the latent heat flux computations, for instance, one year of SSM/I data occupies four (4) Gigabytes. The data set includes each geophysical variable, for each footprint, for every swath for the whole year, organized into month long blocks.

The first step in the SSM/I geophysical data manipulation is transform Level 2 swath data into Level 3 gridded data of fine resolution monthly averaged total precipitable water and surface wind speed (Figure 1). Without GIS tools it is necessary to write our own procedure to perform this task. The Interactive Display Language (IDL) is presently used to perform this tasks. It is likely that a standardized GIS product should be able to perform this sort of function routinely.

The monthly mean SSM/I low level geophysical products, along with AVHRR SST data, are rewritten as ASCII data files which are the inputs to our LHF model written in FORTRAN and run on a Dec alpha. In the case of the SW production scheme, all the initial computational steps are performed on binary data files provided as standard ISCCP products by ISCCP project for 3-hour averages over a seven year period. Once the daily SW fields are calculated IDL is used to produce the monthly averages (Figure 2) and put these in a format which can be combined with the LHF, LW and SE fields. This includes recasting all fields to .5 degree resolution. From this point on, all subsequent operations and visualization are performed in IDL (Figures 3 and 4).

3.2 Statistical and Physical Modeling

Statistical modeling is often used in environmental research to transform low level geophysical products into higher level products. In our application statistical and physical models are both used.

First, the LHF computations require the knowledge of the surface saturation and relative humidity (which are not directly measurable from space). The air mixing ratio, Qa and surface air temperature, Ta, need to be derived from the satellite derived total precipitable water. The goal of that step is to statistically relate the total precipitable water to the surface air humidity (expressed as air mixing ratio, Qa) and temperature. Whereas it has been demonstrated that such a statistical relationship exists over monthly time scale (Liu , 1986), the accuracy of that relationship has been found to vary geographically (Esbensen et al., 1993). A tool that would help in assessing the geographical variations of the accuracy of this relationship and accounting for these variations in the final LHF computations would be very useful.

Second, the LHF computations are performed with a physical model introduced by Liu et al., 1979. This model uses a well known aerodynamic bulk formula (ABF) to relate measurables of moisture, temperature and wind speed to heat and momentum fluxes across the air-sea interface. What makes this model different is the use of similarity theory to determine the transfer coefficients necessary to compute the ABF. Conceptually the similarity theory states that the vertical profiles of moisture, temperature and wind speed should have similar shapes owing to the fact that they are mutually acted upon by the same physical forces throughout space & time. A more detailed theoretical discussion can be found in Businger (1975). The similarity theory requires measurements at two heights, the surface and some common reference height for all three parameters. An iteration technique is then used to find the common profile which has the "best" overall fit for the three parameters.

3.3 Point to Region Relation

The relationship between the total precipitable water (TPW) and the air mixing ratio, Qa, is based on the analysis of radiosonde measurements. Starting from the conceptual notion that Qa will respond to changes in the TPW a statistical relationship was sought between available radiosonde data and TPW. Here we are using information in the vertical direction to modify data in the spatial direction. This method fails for single events, but by averaging over sufficient time scales (~2 weeks) a relationship is found (Liu, 1986). This is a typical example of the application of a statistical relationship for point (radiosonde) data applied to grid data as part of the process of coming up with a high level product. This is probably an area where GIS tools can be useful. In a similar way the total precipitable water is related to the surface air temperature, Ta, which determines the surface saturation humidity.

3.4 Data Blending

A procedure that is often used with satellite data when other data are available to complement them, is the blending of data sets of different origin. This is particularly true when one expects that improved fields can be obtained if the satellite data can be complemented, in some ways, with surface observations in which we have a higher confidence level. Such a data blending approach is used, for instance, to derive global sea surface temperature produced by the Climate Analysis Center of NOAA. SST measured by ships and buoys are used together with AVHRR-derived SST to obtain an optimal SST product for use in climate studies (Reynolds et al., 1988).

The LHF produced by satellite data only (both microwave and infrared measurements) has been found to have some deficiencies in some regions and even systematic errors (Jourdan and Gautier, 1995, Esbensen et al., 1994). We have therefore developed a method for blending satellite-derived products with available in-situ data from the Comprehensive Ocean Atmosphere Data System (COADS) (Slutz et al., 1985) (see Figure 3). The idea being that while the in-situ data may be more accurate, their sampling is very limited and the gradient provided by the satellite data sets can be used to interpolate between the sparse in-situ measurements. This is achieved by assigning weights to the in-situ data sets that depend on the number of measurements used to produce time averages. The satellite data are weighted by the number of samples (nbsTa COADS, nbsQa COADS) used to determine the COADS value. These blended inputs, Qa blend and Ta blend, are then used to recalculate the LHF using the same model as before.

3.5 Spatial Massaging

One of the problems with combining data sets used to compute the four terms that compose the net surface heat flux is that the data sets used have different resolutions. For instance, the ISCCP data and therefore the SW product we compute is only available at 2.5 degree resolution, while the COADS data used to blend the LHF inputs is available at 2 degrees resolution. The method we use to combine these two data sets is to rescale each to the finer resolution of .5 degree resolution and do the arithmetic at this scale. This obviously involves some assumptions regarding the way the interpolation is performed. Here again, GIS tools could be integrated to provide a better way of combining these data sets at different spatial resolution.

3.6 Sensitivity to Assumptions

The blended LHF program uses four inputs: blended air temperature, blended Qa, wind speed and SST. It produces eight output fields: Obukov and Stability Indices, Reynolds Number, latent and sensible drag coefficients, sensible and latent heat fluxes and wind stress. Each of these output fields has a differing sensitivity to the errors/data quality of the model inputs. It would be very useful to have a tool that can assess these sensitivities. This is discussed below.

3.7 Statistical Analysis

In general, some type of statistical analysis is necessary to ensure that the data sets produced are of the highest quality possible. Averages and variance fields are computed to assess whether bias or trends exist. More sophisticated analysis are often performed that are more appropriate to the data set utilized. In our case, for instance, we were interested in investigating the large geographical differences we found between the satellite only fields and the ship only fields. In particular we wanted to investigate the influence of the data sampling on that difference. To address that goal we investigated the correlation between the difference between the two fields and the ship sampling. Considering the difference between the two fields directly results in large areas of strong disagreement, but when an allowance is made for the variable ship sampling of the COADS data, a different picture, shown in figure 5, emerges. We see that when sampling is high (along established shipping lanes) the two fields are more likely to agree than in places where there is little or no ship activity. Satellites have the superior spatial coverage with regions of systematic error, the COADS data is comprised of direct measurements with large spatial gaps, combining these two fields in space is necessary to produce an improved LHF product with the complete coverage needed to estimate oceanic heat transport.

3.8 Typical Procedure Sequence

Here we summarize a typical procedure sequence used to generate high level geophysical products. Some steps may be superfluous or not in the right order, but they are in general representative of those that need to derive such product.

- Swaths -> maps
- Statistical Relationships

    - W -> Qa, W -> Ta
- Physical Modeling
    - (SST, Vs, Qa) -> LHF
    - (Rcloud, Rclear, N) -> SW
- Comparison/evaluation w/ COADS to determine blending, then recalculations using the model
- Straightforward combination of all terms: LHF, SW, LW & SE
- Spatial integration/mathematical operations

4. Visualization

While going through the above steps the user performs a variety of visual confirmations to check the quality and accuracy of the geophysical products obtained at every step. For instance, when the daily maps are derived from the raw SSM/I data, they are viewed in monthly groups to check for missing days, days with large data gaps, and correct registration of each map. At this point a missing step could result in not assigning the correct value to missing data or land areas. This can be quickly checked by viewing the images in monthly groups. Obviously, an automatic procedure that would perform this function flawlessly would speed up the evaluation procedure enormously.

5. GIS Tools Contribution

Some issues where we believe GIS tools could contribute to the development of high level geophysical products such as those described here are:

- Rapid visualization procedure for data quality assessment.
- Combined visualization and statistical analysis package to determine the accuracy of the derived geophysical products.
- Intelligent smoothing procedure on irregularly shaped data such as an ocean basin where there is no data for the land areas.
- Methods to smooth/fill in missing data that may occur over both space and time with varying coherence over the dataset.
- Manage a parallel stream of data quality information along with the product development to be used in sensitivity queries.

6. GIS Unique Functions

At each step in the product development choices are made with regard to data quality: which data to include, how to do smoothing, corrections for known systematic errors etc.. Currently, our final product contains a single frozen collection of these decisions. An area where GIS type tools could be very helpful and have a unique contribution is the management and storage of all these sorts of data quality decisions in order to be able to assess the consequences of altering some combinations of these decisions. With regards to the issues discussed some assessments could be: * What is the impact of including all the available wind speed/water vapor data at the swath level instead of just the ones tagged as "highest quality". * What is the impact of using the various available SST products in determining LHF. (e.g..

MCSST, Pathfinder, OISST) * How do various satellite/solar viewing angle corrections affect our estimate of the heat budgets. * How does the choice of spatial scale or order of operations in computations effect the spatial coherence of the resulting maps. * What is the best way of weighting the number of samples in the COADS data set with SSM/I derived parameters for producing a particular result.

Each of these assessments could also be modified to specific spatial (where the W -> Qa relationship is weak, coastal vs. open ocean, tropical vs. southern ocean) or temporal (El Nino vs. non-El Nino years) considerations.

In this way, the reporting phase of our scientific result could be enhanced by showing appropriate error ranges that are a function of the spatial averaging processes or data selection for instance.

The idea would be to store the influences of propagation all the data quality decisions as a parallel process to the geophysical product development and come up with a comprehensive data quality description for the final product. This comprehensive data quality description would be the basis for answering the above type product assessments.

7. Acknowledgments

This study has been made possible by a grant from Jet Propulsion Laboratory (JPL 959177). We would also like to thank Bill O'Hirok for his useful suggestions and James Marquez for helping with the figures and manuscript preparation.

8. Bibliography

Businger, J.A., 1975: Interaction of sea and atmosphere. *Rev. Geophys. Space Phys.*, 13, 720-822.

Esbensen, S.K.,D.B. Chelton, D. Vickers and J.Sun, 1993: An analysis of errors in Special Sensor Microwave Imager evaporation estimates over the global oceans. *Geophys. Res.*, 98, 7081-7101.

Gautier, C., G. Diak and S. Masse, 1980: A Simple physical model to estimate incident solar radiation at the surface from GOES satellite data. *J. Appl. Meteor.*, 19(8), 1005-1012

Jourdan, D., and C. Gautier, 1995: Comparison between latent heat flux computed from multisensor (SSM/I and AVHRR) and from in-situ data. *J. Atmos. Ocean. Technol.*, 12, 46-72

Liu, W.T., K.B. Katsaros and J.A. Businger, 1979: Bulk parameterization of air-sea exchanges of heat and water vapor including the molecular constraints at the surface. *J. Atm. Sci.*, Vol. 36, 1722-1735.

Liu, W.T., 1986: Statistical relation between monthly mean precipitable water and

surface-level humidity over global oceans. *Mon. Wea. Rev.*, 114(8), 1591-1602.

Reynolds, R.W.. 1988: A real-time global sea surface temperature. . *J. Climate*, 1, 75-86

Rossow, W.B., L.C. Gardev, P-J Lu and A.W. Walker, 1991: International Satellite Cloud Climatology Project (ISCCP) Documentation of cloud data. WMO/TD-No. 266, World Meteo. Org., Geneva.

Smith, E.A., 1992: User's guide to the NOAA Advanced Very High Resolution Radiometer MCSST dataset. JPL D-10737, 14 pp.

Wentz, F.J., 1983: A model function for ocean microwave brightness temperature. *J. Geophys. Res.*, 88 (C3), 1892-1908.

Wentz, F.J., 1992: User's Manual SSM/I Monthly Ocean Tape, Remote Sensing Systems Tech. Rep.

Catherine Gautier and Pete Peterson
Earth-Space Research Group
Institute for Computational Earth System Science
University of California Santa Barbara
Santa Barbara, CA, 93106-3060
Phone: (805) 893-4912, Fax: (805) 893-2578
e-mail: pete@icess.ucsb.edu

Limin Yang
Zhi-Liang Zhu
Jorge A Izaurralde
James W. Merchant

Evaluation of North and South America AVHRR 1-km Data for global environmental modeling

Abstract

A new time series of satellite data has been produced for the continents of North and South America based on specifications developed through the International Geosphere Biosphere Programme Data and Information System (IGBP-DIS). The data provide, for the first time, complete and consistent coverage over the western hemisphere land area at the full 1 km resolution of the Advanced Very High Resolution Radiometer (AVHRR) sensor. This data set and the derived land cover characteristics product currently being developed will have many applications for global environmental modeling research.

An evaluation of North and South America AVHRR data sets documents its salient characteristics. The analysis focused on scan angle distribution, image area distortion as a result of map projection, the effect of view geometry on normalized difference vegetation index (NDVI) compositing, the distribution of high solar zenith angle, and cloud contamination in NDVI composites. The evaluation shows that, for North America data, the compositing procedure results in a bias favoring off-nadir pixels, particularly pixels at post-nadir (forward scanning) positions in the winter months. In contrast, such a bias does not exist in the South America counterpart. Image area distortion due to reprojection from image to map and data resampling procedure have been quantified. The results for scan angle distribution and image area distortion provide a basis for calculating the effective minimum mapping area for various geographic locations. The amount of missing data caused by large solar zenith angle ranges from 0 to 42 percent, and the estimated percentage of cloud residuals in the 10-day composites varies from 2 to 10 percent depending on seasons. Suggestions are made with regard to handling missing data and reducing cloud effect through NDVI recompositing.

---

1. Requirement for global land data set for environmental modeling

Reliable, timely, and consistent global land data are essential for modeling and monitoring Earth's environment (Townshend 1992, Sellers et al. 1986, Pielke 1991, Townshend 1994, Dickinson 1995). The exchange of momentum, heat, moisture and carbon dioxide across the land-atmosphere interface affects the dynamics of the Earth system. Land cover characteristics and land cover conversion also significantly influence global and regional biogeochemical cycles of carbon, nitrogen, and other elements (Townshend, 1994). Thus, for climate, hydrological, and ecological modeling, detailed data on land characteristics (e.g. land cover, surface biophysical parameters, vegetation dynamics, topography, and soils) at various spatial and temporal scales are required.

2. Remote sensing data for global environmental modeling and monitoring

Remote sensing techniques are the only practical method to acquire information on the characteristics and dynamics of global land surface. Multi-temporal and multi-spatial resolution data obtained via remote sensing techniques enable the derivation of terrain characteristics at various scales and allow quantitative description of land cover dynamics (e.g. surface albedo, aerodynamic roughness, canopy conductance, and vegetation condition). The challenge for remote sensing, however, is whether it can reliably and consistently provide such needed data to meet specific requirements of global change research.

During the past decade, one type of data most widely used for global and regional environmental modeling and monitoring is that acquired by the National Atmospheric and Oceanic Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR). The advantages and limitations of AVHRR data have been well documented (Holben 1986, Goward et al. 1991, Los et al. 1994). The main advantages of the AVHRR data include:

- nearly daily global coverage useful for study of land cover and vegetation dynamics,
- acceptable spatial resolution and manageable data volume for global or continental environmental studies,
- continued data acquisition since early 1980,
- inexpensive and ease of data accessibility.

Major uncertainties and limitations are:

- lack of on-board calibration for channel 1 and channel 2,
- the drifting of the satellite orbit and degradation of sensor performance through time,
- geometric distortion due to wide scan angles and variations in sun-sensor-target geometry,
- effect of cloud and atmospheric attenuation on recorded signals,
- effect of land surface bi-directional reflectance properties on recorded signals.

Despite the inherent limitations, AVHRR data have been widely used in land cover studies and are one of the important data sources for global change research. The International

Geosphere-Biosphere Programme (IGBP) has identified AVHRR as the prime data source for global land cover mapping. A detailed summary of currently available global land data sets developed using AVHRR is presented by Townshend (1994).

3. Development of the global 1 km AVHRR land data

To facilitate global change research, the IGBP promotes development of a 1-km AVHRR global land data set which, for the first time, will provide a complete coverage of the global land surface at the full AVHRR resolution. The scientific requirements for such global data sets have been identified by several organizations (Townshend 1992, Malingreau and Belward 1994, Running et al. 1994). Currently, the data implementation has been conducted at the U.S. Geological Survey (USGS), Earth Resources Observation Systems (EROS) Data Center (EDC) as a part of its mission as the National Aeronautics and Space Administration (NASA) Earth Observation System (EOS) Land Processes Distributed Active Archive Center (LPDAAC). The primary sponsors of this project are the USGS, NASA, NOAA, and European Space Agency (ESA) (Eidenshink and Faundeen, 1994).

The global 1 km AVHRR data processing standards are developed under the guidance of the IBGP (Townshend 1992). Major data processing procedures include data calibration, atmospheric correction, geometric registration, and global 10-day maximum normalized difference vegetation index (NDVI) composites. The resulting composites represent a 10-day (3 per month) time series and contain the following 10 bands:

- AVHRR channels 1-5,
- NDVI,
- satellite zenith,
- solar zenith,
- relative solar/satellite azimuth,
- date index of pixel observation.

The production of the North and South America 1-km data set has been completed at the EDC. This data set consists of five Goode Interrupted Homolosine projection regions and thirty-six 10-day composites from April 1-10, 1992 (first period) to March 21-31, 1993 (last period). The North and South America data are accessible through the USGS EROS Data Center home page under EOS LPDAAC section.

4. Data quality evaluation and documentation

It has been recommended that remote sensing data made available to general users be fully documented (Justice and Townshend, 1994). For North and South America 1-km AVHRR data, an evaluation of the NDVI composites has been undertaken in order to document and understand their salient qualities. The evaluation focused on (1) determining the distribution of sensor scan angle in the composite data, (2) evaluating image area distortions due to the use of the Goode projection and resampling technique, (3) documenting high solar zenith angle effect, and (4) assessing the presence of residual clouds in the data set. The following section documents the primary characteristics of the data set followed by recommendations to potential data users.

5. Characteristics of North/South America 1 km NDVI Composites

5.1 Viewing geometry

One important characteristic of the data set is its distribution of viewing geometry. Compared to other sensor, AVHRR has a wide field of view which, when coupled with Earth's curvature, causes geometric and spectral distortions for pixels at off-nadir positions. Previous work has shown that NDVI, and, subsequently, the compositing process, can be affected by changes in AVHRR scan angle (Holben and Fraser 1984, Deering and Eck 1987, Gutman 1991, Moody and Strahler 1994), and that using atmospherically corrected data in compositing may shift pixel selection toward high off-nadir positions (Cihlar and Huang 1994). In the current study, the per pixel scan angle values were derived from satellite zenith band in or near the solar principal plane (95 percent of the pixels are within +/- 40 degree relative azimuth angle). Angle values range from 0 to 50 degrees for both pre-nadir (backscatter) and post-nadir (forescatter) scanning directions.

The viewing geometry was investigated for the overall weighted-average scan angle and for latitudinal and longitudinal effects. Results show that, on average, compositing bias toward post-nadir direction exists for North America data (figure 1). This is consistent with work previously reported (Gutman 1991, Moody and Strahler 1994). Most of the bias may be attributed to preferential selection of post-nadir pixels in the winter months (figure 2), during which a temporal variation of the mean scan angle is evident. In South America, however, the frequency distribution of scan angle is much more uniform and symmetrical around nadir (figure 1), and the weighted mean scan angle is also close to nadir (figure 2) - a significant improvement over the apparent temporal bias shown in the North America data.

The suspected effects due to locations of satellite orbits relative to ground receiving stations were investigated by examining any latitude or longitude biases. It is noted that no apparent bias exists due to differences in latitude over the two continents. The longitude bias may exist in some cases, but it does not occur consistently at all latitudes or locations. Further investigation in this area is warranted.

5.2 Image area distortions due to map projection

Although the use of the Goode projection reduces image distortions in large land areas, changes in local scale and resolution of the original image still occur as the result of data being compressed and expanded to fit the map projection. To examine these image distortions, a technique reported by Steinwand et al. (1995) was adopted. A checkerboard image, consisting of regularly spaced 10-by-10 pixel squares, was created; each square had pixel values ranging from 1 to 100. This checkerboard image was reprojected to various regions of the Goode projection, using georegistration grids and the nearest neighbor resampling method. Pixel values of the squares (now skewed) were counted in the new map projection space. Changes in these values represent image area distortions caused by both the use of map projection and the result of the resampling method (Steinwand et al. 1995).

In this study, sample windows were taken from representative geographic locations ranging from 40.19 S to 84.80 N latitude. As indicated in Table 1, image area expansion occurs at most sample locations, with maximum duplication up to five times as high. Image area reduction also occurs, but not as often as expansion. Data losses due to area reduction range from 0 to 31 percent, resulting in changes in the spatial resolution of the original data set. The observed distortions are related to individual satellite orbit and projection grids. There is no consistent pattern associated with locations of the sample windows. Note that a portion of the image area distortions may be related to the use of nearest neighbor resampling. The error may be reduced if cubic convolution or other interpolation resampling techniques are used.

The analyses of viewing geometry and image area distortion indicate that the 1.1 km AVHRR nominal pixel scale (at nadir) is modified by the two variables. An effective minimum mapping unit may be determined for a given geographic area based on these variables, and the information is useful for land cover mapping applications.

5.3 Solar zenith angle effect

Missing data exist in the North/South America 1-km data as a result of masking for large solar zenith angle (SZA). The low solar illumination and long atmospheric path length encountered at the high latitudes during the low-sun seasons seriously degrade the data quality in representing land cover features. Due to this limitation, all pixels that have SZA greater than 80 degrees are "flagged" in maximum NDVI composite processing; no NDVI is computed for these pixels.

Based on spherical trigonometry, the SZA is a function of the latitude of the location, the solar declination, and the time of day. Thus, the areal extent of the missing AVHRR data due to cut-off SZA is a function of geographic location and the time of the year, large in the low sun seasons and small in the high-sun seasons. Table 2 lists statistics of missing data for all 12 monthly composites from April 1992 to March 1993. Note that, for North America, the percentage of missing data increases from 1 percent of total pixels in July to 42 percent in January; on the other hand, missing data in South America are found only in a period from March through September, and the highest percentage is only 1.1 percent in July. The significantly low amount of missing data in South America is apparently due to confined land area at higher latitudes.

5.4 Assessing cloud contamination in NDVI composites

Even though the maximum value composite (MVC) technique reduces cloud effects, the resulting composite data are usually not completely cloud-free. In areas of persistent cloudiness (e.g. tropical latitudes and some coastal areas) or frequent occurrence of subpixel sized clouds, it is difficult to remove cloud contamination through the maximum NDVI compositing method (Goward et al. 1991).

The assessment for cloud contamination of the North and South America NDVI composites was conducted using a 1-percent systematic sample of the original data; the cloud-screening method was adopted from several studies for assessing global and continental distribution of cloud cover using AVHRR data (See Baglio and Holroyd 1989, Stowe et al. 1991 for more details). The method utilized AVHRR channels 1 through 5, along with solar illumination and viewing geometry for separating clouds from other land features. In this study, three cloud screening tests were applied to each of the 10-day composites. The first test was based on the fact that most land surfaces have much lower reflectance than clouds do in AVHRR channel 1. The second test used channel 4 minus channel 5 radiative temperature difference to detect thin cirrus clouds and clouds in polar latitude (Stowe et al. 1991). The third test used the temperature difference between channel 3 and channel 4 to separate cloud from cold, snow-covered land (Baglio and Holroyd 1989).

Table 3 summarizes estimated cloud residuals for the 10-day composites. For North America data, the amount of cloud pixels varies from less than 2 percent of total land pixels in November to almost 8 percent in May. The 10-day composites of spring (April and May), midsummer (August), and fall (late September to early October) exhibit higher percentage of cloud cover than those of winter months (November through January). In comparison, South America data show a higher percentage of cloud. The amount of cloud pixels varies from 4.6 percent in July to 10 percent in January. The most cloud-prone regions are found in the tropical latitudes of both hemispheres (especially in the Amazon Basin), and areas of Labrador east of Hudson Bay, Canada as well as other coastal regions. Because no attempt was made in this study to identify subpixel sized clouds, the cloud area estimates from this study are rather conservative. Figure 3 and 4 illustrate, as an example, the spatial distribution of clouds identified from the April 21-30 composite of North America and the January 1-10 composite of South America as well as corresponding AVHRR channel 1, 4, and NDVI.

An experiment was conducted to assess potential reduction of cloud attenuation through recompositing of NDVI time-series over a 30-day period (versus 10-day period). The results from this experiment indicate that recompositing over a longer period can further reduce cloud attenuated pixels. For example, the amount of cloud-affected pixels identified from the North America data ranges from 3.7 to 6.2 percent for the 3 August 10-day composite and 3.2 to 6.4 percent for the 3 October composites; the August and October monthly composites contain only 1.3 and 3.1 percent estimated cloud pixels, respectively.

6. Potential high level product from global 1 km AVHRR data for global and regional environmental modeling and monitoring

One of the primary usages of the global 1 km land data is to develop a global land cover data base, which has been endorsed by the IGBP land cover core project (Townshend 1992). This project will utilize the multi-temporal NDVI data, in conjunction with various ancillary data such as digital elevation model, ecoregion and climate, to derive crucial information on land cover for the entire global land mass at the full AVHRR resolution. The development of the land cover data follows a concept of flexible land cover characteristics database previously developed (Loveland et al. 1991, Brown et al. 1993, Reed et al. 1994), which allows the user to tailor the database for their specific needs for various environmental modeling studies.

The proposed global land cover characteristics database will include seasonal land cover regions with relatively homogeneous land cover associations and distinct phenology (onset, peak and duration of greenness). The database will also contain other spatial data which are utilized to develop and characterize the land cover regions. Those data sets include both remotely sensed data and other ancillary data (e.g. topography, ecoregions, land cover/use, and climate). The quantitative and qualitative descriptive attributes for each of the land cover regions are another component of the database. These attributes serve as indicators of the related land cover regions, and allow a translation from the seasonal land cover regions to other commonly used land cover schemes, such as the USGS Anderson land cover type (Anderson et al. 1976), the Biosphere-Atmosphere Transfer Scheme (Dickinson et al. 1986), the Simple Biosphere model (Sellers et al. 1986) and others.

Scientists at the EDC and the University of Nebraska-Lincoln are now developing the North and South America land cover characteristics data base. The preliminary results from this effort are being summarized, and the final land cover data base will be made available to the user communities.

7. Summary and Conclusions

A new satellite-derived time series of North and South America land data have been produced by USGS EROS Data Center based on specifications developed through the International Geosphere Biosphere Programme Data and Information System (IGBP- DIS). The data consist of thirty six AVHRR NDVI 10-day composites, all five spectral bands and viewing geometry over a 12 month period from April 1992 to March 1993. This data set provides, for the first time, complete and consistent coverage over the western hemisphere land area at the full 1-km resolution of the AVHRR sensor. In addition, one of the primary high level products to be derived from these data is a new global land cover characteristics database with the same spatial resolution. These data sets have many potential applications for global environmental modeling research.

Evaluation of the North and South America AVHRR data was conducted and the key characteristics of the data set are:

- The compositing procedure selects off-nadir pixels in the North America data set, despite the absence of atmospheric correction prior to compositing. In contrast, the South America data do not have such a scan angle bias.
- The global Goode projection used in the data set has data compression and duplication effects, which cause changes in the spatial resolution. The effects vary with locations and can be measured using a simple "checkerboard" approach.
- The 1.1-km nominal pixel size (at nadir) is modified by variables such as the high scan angle and image area distortion described. An effective minimum mapping area can be determined for a given geographic area based on these variables.
- Due to high solar zenith angle (SZA) effect missing data exist in the North and South America NDVI composites. The areal extent of the missing data ranges from 0 to 42 percent over the 12 month period. For deriving land cover information, a best estimate for winter missing data is adequate considering the area is relatively homogeneous in ground cover with low vegetation but lasting snow and ice. However, for deriving

surface biophysical parameters from the data (for example, albedo and leaf area index), temporal or spatial interpolation, or both methods are preferable especially when additional information on surface condition is available (for example, Landsat data or ground measurements). Caution must be taken, however, to minimize potential distortion of the original data due to limitations of the interpolation algorithms.

- The estimated percentage of cloud pixels ranges from 2 percent to 11 percent. The percentage is relatively high in warm seasons and low in the winter, and is higher in the South America data. The effect of cloud contamination may be reduced by extending the composite period if no significant compromise has to be made by the users. The trade-off between extending the NDVI compositing period and lowering temporal resolution, as well as potential propagation of spatial registration errors, needs to be assessed based on research objectives and geographic area. For regional or local applications, it is recommended that the users examine each 10-day composite for their specific study area, as the quality of 10-day composites varies with geographic locations and time. In this study, comparison of 10- versus 30- day composites suggests that recompositing over a longer period (for example, monthly) is preferable for large area land cover mapping.

Acknowledgments

The authors would like to thank David Meyer and Bradley Reed of the EROS Data Center for reviewing the manuscript and comments for its improvement. Limin Yang and James Merchant's work was supported through U.S. National Aeronautics and Space Administration Grant (NAGW- 3940). Zhi-Liang Zhu's work was performed under USGS contract 1434-92-C-40004.

End Notes

This manuscript is currently undergoing formal approval by the United States Geological Survey.

References

- ANDERSON, J.R., HARDY, E.E., ROACH, J.T. and WITMER, R.E. (1976) A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey professional paper 964.

- BAGLIO, J.V., and HOLROYD, E.W. III (1989) Methods for operational snow cover area mapping using the Advanced Very High Resolution Radiometer: San Juan Mountains test study. USGS Research Technical Report, USGS/EROS Data Center, Sioux Falls, SD.

- BROWN, J.F., LOVELAND, T.R., MERCHANT, J.W., REED, B.C., and OHLEN, D.O. (1993) Using multisource data in global land characterization: concepts, requirements, and methods. Photogrammetric Engineering & Remote Sensing, 59, 977-

987.

- CIHLAR, J., and HUANG, F. (1994) Effects of atmospheric correction and viewing and restriction on AVHRR data composites. Canadian Journal of Remote Sensing, 20, 132-137.
- DEERING, D.W., and ECK, T.F. (1987) Atmospheric optical depth effects on angular anisotropy of plant canopy reflectance. International Journal of Remote Sensing, 8, 893-916.
- DICKINSON, R.E., Henderson-Sellers, A., Kennedy, P.J., and Wilson, M.F. (1986) Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model. NCAR Tech Note 275+STR. National Center for Atmospheric Research, Boulder, Colorado.
- DICKINSON, R.E. (1995) Land processes in climate models. Remote Sensing of Environment, 51, 27-28.
- EIDENSHINK, J.C., and FAUNDEEN, J.L. (1994) The 1-km AVHRR global land data set: first stages in implementation. International Journal of Remote Sensing, 15, 3443-3462.
- GOWARD, S.N., MARKHAM, B, DYE, D.G., DULANEY, W., and YANG, J. (1991) Normalized difference vegetation index measurements from the Advanced Very High Resolution Radiometer. Remote Sensing of Environment, 35, 257-277.
- GUTMAN, G.G. (1991) Vegetation indices from AVHRR: an update and future prospects. Remote Sensing of Environment, 35, 121- 136.
- HOLBEN, B.N. (1986) Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing, 7, 1417-1434.
- HOLBEN, B.N., and FRASER, R.S. (1984) Red and near-infrared sensor response to off-nadir viewing. International Journal of Remote Sensing, 5, 145-160.
- JUSTICE, C.O., and TOWNSHEND, J.R.G. (1994) Data sets for global remote sensing: lessons learnt. International Journal of Remote Sensing, 15, 3621-3639.
- LOS, S.O., JUSTICE, C.O., and TUCKER, C.J. (1994) A global 1 by 1 degree NDVI data set for climate studies derived from the GIMMS continental NDVI data. International Journal of Remote Sensing, 15, 3493-3518.
- LOVELAND, T.R., MERCHANT, J.W., OHLEN, D.O., and BROWN, J.F. (1991) Development of a land cover characteristics data base for the conterminous U.S. Photogrammetric Engineering & Remote Sensing. 57, 1453-1463.
- MALINGREAU, J.-P., and BELWARD, A.S. (1994) Recent activities in the European Community for the creation and analysis of global AVHRR data sets. International Journal of Remote Sensing, 15, 3397-3416.

- MOODY, A., and STRAHLER, A.H. (1994) Characteristics of composited AVHRR data and problems in their classification. International Journal of Remote Sensing, 15, 3473-3491.

- PIELKE, R.A., DALU, G.A., SNOOK, J.S., LEE, T,J., and KITTEL, T.G.F. (1991) Nonlinear influence of mesoscale land use on weather and climate. Journal of Climatology, 4, 1053- 1069.

- REED, B.C., LOVELAND, T.R., STEYAERT, L.T., BROWN, J.F., MERCHANT, J.W., and OHLEN, D.O. (1994) Design global land cover databases to maximize utility, in Environmental Information Management and Analysis: Ecosystem to Global Scales. W.K.MICHENER, J.W.BRUNT, and S.G.STAFFORD, editors: London, Francis and Taylor, 299-314.

- RUNNING, S.W., JUSTICE, C.O., SALOMONSON, V.V., HALL, D., BARKER, J., KAUFMANN, Y.J., STRAHLER, A.H., HUETE, A.R., MULLER, J-P., VANDERBILT, V., WAN, Z.M., TEILLET, P., and CARNEGGIE, D. (1994) Terrestrial remote sensing science and algorithms planned for the Moderate Resolution Imaging Spectrometer (MODIS) of the Earth Observing System (EOS). International Journal of Remote Sensing, 15, 3587-3620.

- SELLERS, P.J., MINTZ, Y., SUD, Y.C., and DALCHER, A. (1986) A simple biosphere model (SiB) for use with general circulation models. Journal of Atmospheric Sciences, 43, 505-531.

- STEINWAND, D.R. (1994) Mapping raster imagery to the Interrupted Goode Homolosine projection. International Journal of Remote Sensing, 15, 3463-3471.

- STEINWAND, D.R., HUTCHINSON, J.A., and SNYDER, J.P. (1995) Map projections for global and continental data sets and an analysis of pixel distortion caused by reprojection. Photogrammetric Engineering & Remote Sensing, 61, 1487-1497.

- STOWE, L.L., McCLAIN, E.P., CAREY, R., GUTMAN, G.G., DAVIS, P., LONG, C., and HART, S (1991) Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. Advance in Space Research, 11, 351-354.

- TOWNSHEND, J.R.G. ed. (1992) The global 1 km AVHRR data set: further recommendations. IGBP-DIS Working Paper No. 3. The International Geosphere Biosphere Programme Data and Information System, University of Maryland, College Park, MD, U.S.A.

- TOWNSHEND, J.R.G. (1994) Global data sets for land applications from the Advanced Very High Resolution Radiometer: an introduction. International Journal of Remote Sensing, 15, 3319-3332.

Figure 1. Distribution of scan angles summarized for the North and South America data. Scan angle data are averaged for all 36 composite periods.

Figure 2. Weighted mean scan angle summarized for the North and South America data for the 36 composites. Pre-nadir as negative value and post-nadir as positive value.

Figure 3. Estimated cloud distribution in the North America data set, April 21-30, 1992.

Figure 4. Estimated cloud distribution in the South America data set, January 1-10, 1993.

Table 1. Image area distortions (changes in number of pixels) due to the use of Interrupted Goode Homolosine projection and resampling method.

Sample center latitude	Sample center longitude	Original data area used (%)	Pixel duplication average	Pixel duplication maximum
North America				
20.85N	88.08W	100	1.93	3
24.45N	101.53W	99	1.16	2
31.01N	98.99W	81	1.11	2
39.74N	91.20W	100	2.10	3
40.01N	112.98W	100	2.18	3
41.76N	81.69W	100	2.24	4
41.88N	61.13W	100	1.65	2
42.62N	63.27W	100	2.05	4
44.85N	84.35W	100	1.64	3
45.74N	163.35W	100	2.01	3
48.13N	140.79W	100	2.09	3
52.93N	97.31W	88	1.00	1
59.13N	82.00W	88	1.00	1
59.86N	159.54W	89	1.00	1
59.96N	61.48W	89	1.00	1
60.71N	179.06W	100	2.12	3
63.78N	125.76W	100	1.96	3
64.73N	108.24W	100	2.02	3
68.97N	86.81W	88	2.60	3
71.15N	149.83E	100	2.17	4
71.32N	69.69W	85	2.48	3
74.68N	129.55W	100	1.60	2
82.44N	157.88W	71	2.68	4
84.80N	85.16W	69	2.17	4
South America				
9.93N	74.02W	100	1.57	3
5.08N	59.78W	100	1.87	4
0.13N	66.24W	88	0.96	2
6.83S	63.50W	94	1.02	2
19.66S	45.39W	100	2.32	3
25.84S	63.89W	100	1.84	4
32.35S	54.13W	84	0.91	2
40.19S	65.58W	100	1.94	5

Table 2. Amount of missing data due to solar zenith angle greater than 80 degrees for 12 NDVI monthly composites.

Month	Year	Number of pixels	Percent

North America			
Apr	1992	1,333,346	5.2
May	1992	1,212,581	4.7
Jun	1992	947,088	3.7
Jul	1992	254,128	1.0
Aug	1992	1,447,545	5.7
Sep	1992	985,699	3.8
Oct	1992	2,448,994	9.6
Nov	1992	7,616,120	30.0
Dec	1992	10,732,749	42.3
Jan	1993	8,027,949	31.7
Feb	1993	3,949,743	15.5
Mar	1993	1,952,302	7.7
South America			
Apr	1992	0	0
May	1992	35,330	0.1
Jun	1992	2,812	0.0
Jul	1992	225,020	1.0
Aug	1992	6,665	0.0
Sep	1992	4,153	0.0
Oct	1992	1	0.0
Nov	1992	0	0.0
Dec	1992	0	0.0
Jan	1993	0	0.0
Feb	1993	0	0.0
Mar	1993	623	0.0

Table 3. Amount of cloud pixels estimated for 10-day composites, April 1992 to March 1993. The estimates are based on an 1-percent sample of the original composite data.

Composite Period		Percent	Composite Period		Percent
North America					
Apr	01-10 1992	6.7	Oct	01-10 1992	6.4
Apr	11-20 1992	5.3	Oct	11-20 1992	3.2
Apr	21-30 1992	6.6	Oct	21-30 1992	4.0
May	01-10 1992	6.3	Nov	01-10 1992	2.9
May	11-20 1992	3.7	Nov	11-20 1992	2.4
May	21-30 1992	7.5	Nov	21-30 1992	1.6
Jun	01-10 1992	4.2	Dec	01-10 1992	1.9
Jun	11-20 1992	4.6	Dec	11-20 1992	3.2
Jun	21-30 1992	3.0	Dec	21-30 1992	3.1
Jul	01-10 1992	4.7	Jan	01-10 1993	3.6
Jul	11-20 1992	3.5	Jan	11-20 1993	3.1
Jul	21-30 1992	3.9	Jan	21-30 1993	2.4
Aug	01-10 1992	3.7	Feb	01-10 1993	2.3
Aug	11-20 1992	6.2	Feb	11-20 1993	4.4
Aug	21-30 1992	6.0	Feb	21-28 1993	3.7
Sep	01-10 1992	3.0	Mar	01-10 1993	4.5
Sep	11-20 1992	3.7	Mar	11-20 1993	2.7
Sep	21-30 1992	5.7	Mar	21-30 1993	4.1
South America					
Apr	01-10 1992	4.9	Oct	01-10 1992	5.3

Jul	01-10 1992	4.6	Jan	01-10 1993	10.1

Authors

Limin Yang
Center for Advanced Land Management Information Technologies
Conservation and Survey Division
University of Nebraska-Lincoln
Lincoln, Nebraska 68588-0517
U.S.A.
Telephone: (605) 594-6039
Fax: (605) 594-6589
Email: lyang@unl.cr.usgs.gov

Zhi-Liang Zhu
Hughes STX
US Geological Survey
EROS Data Center
Sioux Falls, South Dakota 57198
U.S.A.

Jorge A Izaurralde
Faculty of Agronomy
National University of Cordoba
Cordoba, ARGENTINA

James W. Merchant
Center for Advanced Land Management Information Technologies
Conservation and Survey Division
University of Nebraska-Lincoln
Lincoln, Nebraska 68588-0517
U.S.A.

D.C.L.Lam, D.A. Swayne, C.I. Mayfield, D.D. Cowan

INTEGRATION OF GIS WITH OTHER SOFTWARE SYSTEMS: Integration versus Interconnection

---

There are very difficult practical issues of systems integration in environmental modeling, particularly when data collections are large and the systems being modeled are complex.

GIS, environmental modeling and statistical data analysis are three examples of computational activities which are usually relegated to separate, large and sophisticated computer systems. Each has (usually) considerably more functionality than is required by a single user whose needs exceed the range of one of them alone (e.g. both GIS and statistical data analysis). We will examine issues such as: tools for maintaining commonality in "look and feel", data transfer between processing domains (interoperability vs. efficiency), architectures (e.g. client-server) to bridge between applications.

The purpose of this paper is to bring our observations to the attention of users who want to view a collection of software tools as a single system and software developers who recognize the problems of multiple software systems integration. We would hope to engage the attention of researchers who have attempted solutions such as software repartitioning, use of visual programming and object-oriented interface development tools, particularly for large databases.

In our experience, data organization benefits from a GIS perspective. Very simple issues like the quality coverage of the data in space can be resolved by a simple geographical display. Consequently, why not use a simple pictorial geographical query? Likewise, the quality of the time-dimension over a geographical region can be combined with a geographical display by visualizing (contouring or depth shading).

For these two examples (spatial and temporal coverage of data), for interfacing models over a large range of possible input parameters, and for many other circumstances, it is beneficial to have a level of integration which is much greater than simple file transfer mode. Shared link libraries, similar styles of presentation (e.g. spreadsheet or table), and similarity of function all contribute to successful integration.

Various aspects of our joint work at National Water Research Institute, University of Waterloo and University of Guelph have contributed to our perceptions of the integration problem, including data analyses, modeling interfaces and map-based queries on large databases.

We envision an eventual environment in which a toolkit is presented, for which components interconnect smoothly, and for which a user selects only those tool groupings which are needed for the task at hand, and for which external applications (such as engineering models) can be attached in an orderly fashion by an end user.

Dr. Lam is a Senior Scientist, New Technologies Research Branch, National Water Research Institute, Environment Canada, 867 Lakeshore Road Burlington, Ontario, L7R 4A6 CANADA David.Lam@cciw.ca

---

Dr. Swayne is a Professor of Computing and Information Science, University of Guelph, Guelph, NIG 2W1, Ontario. dswayne@uoguelph.ca

---

Dr. Cowan is Professor of Computer Science University of Waterloo, Waterloo, N2L 3G1, Ontario. DCowan@WatCSG.UWaterloo.ca

---

Dr. Mayfield is Professor of Biology University of Waterloo, Waterloo, N2L 3G1 mayfield@sciborg.uwaterloo.ca

---

Introduction -Dimensions of the Problem

Environmental modeling software is often rather large in scope and necessitates a considerable investment in learning and in computational resources. Software such as SWMM4 for urban watersheds or AGNPS for rural ones typically comes with a very large data input requirement. Within a research group, one or two team members will become familiar with the requirements of the modeling system, while others concentrate on other aspects of the modeling enterprise (e.g. GIS, data assembly and verification).

Inputs from air quality simulation can be even more demanding of resources, since air quality models are typically equivalent in horizontal scope, but with several layers in the vertical direction to be managed. Such demands often require a person or even a team to concentrate on very large data collections representing wind vector files, atmospheric chemistry data and simulation.

The Role of GIS

A natural playing field for both the inputs to, and the outputs from, models which have been assembled to solve a problem is the geographical information system (GIS).

1. A GIS is certainly required to visualize the results.
2. A GIS is extremely helpful in setting up the models, but in order to facilitate model development,
3. a GIS is a natural query language for facts to be input to models, and for comparison of models to data.

It has been our experience that GIS "packages" are extremely complex entities. They function as repositories for the map layers for which they are deservedly well-organized to manipulate.

They do not, however, provide good interactive (or automatic) point-by-point query as a front end to a time-series database, for instance.

They are not configured for interaction with complex modeling software. That software often comes with its own visualization (or, more frequently, no visualization). Data import / export is discouraged (and puts programmers in the same position of power as the draughtsmen and cartographers of old.

In contrast to what is offered, the functionality the GIS user requires is often very rudimentary. There are, roughly speaking, five steps in the modelling process (ignoring interior loops and dead-ends). There is a clear role for GIS in each of them, but in some plases, the tool is too specific, and in others it is not simple enough.

The phases we identify are:

1. exploration phase
2. tool assembly phase
3. trial and verification phase
4. production and verification phase
5. report / presentation generation phase

The phase most suited to GIS is the number 5 - report / presentation generation. We will explore our suggestions for the remaining four and indicate some possible solutions for the problems we have identified with the current situation.

Exploration Phase - Data and Model Exploration

For environmental modeling, it is important to see where the data are located within a region of interest. Sampling plans must include coverage of interesting or important zones. They must also include suitable background or reference locations. If soils are important, then sampling must be planned to intersect with various soil types. Air quality monitoring within a region (even a country) must include boundary points external to the region. These data and model points are better served when spatial information is easily accessible to the modeler.

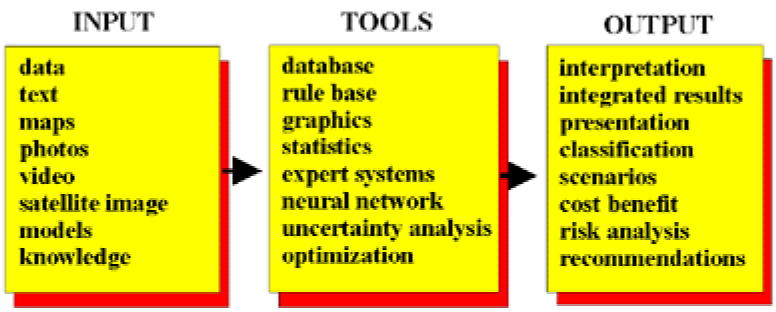


Figure 1. The RAISON Decision Support System

Database query founded on neighborhood relationships (within polygon, associated to object, etc.) is a fundamental requirement for data exploration. Our group has begun using a tool developed within one of our projects, MapMosaic, for spatial queries.

The background conditions for modeling can also be pinpointed with GIS. For this part of the exploration phase, more background information about the models must be made available. A consistent on-line help facility for the models, and a quality explanation facility for the model interface must be made available. Input parameters (and even the units of measurement) change radically, depending on the models used. Flexibility and ease of use increase the quality of the results obtained.

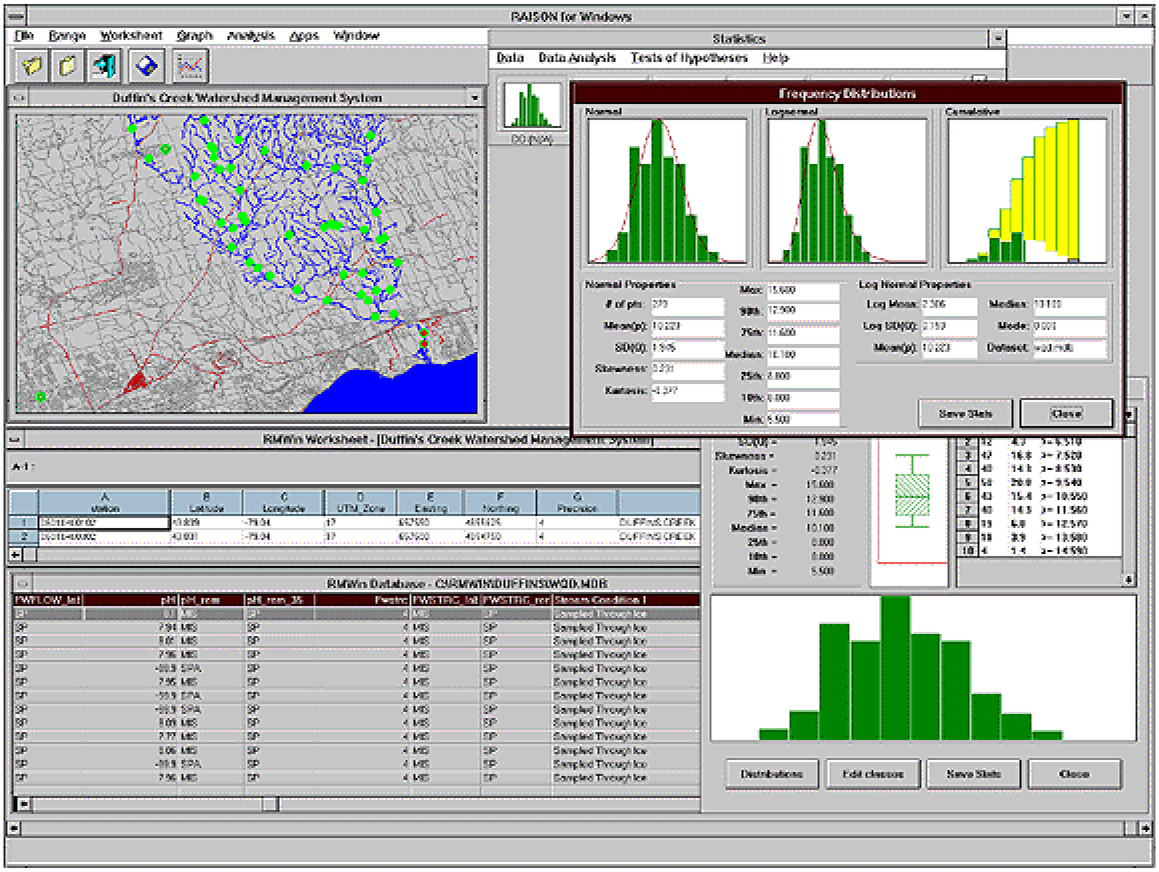


Figure 2. Database View

The flexibility of combining different input would also greatly facilitate the appropriate selection of tools for generating desirable output. An example is the framework adopted in the RAISON system developed at Environment Canada (Lam and Swayne, 1991). As shown in Figure 1, this system is designed to absorb or linked to various types of input, including numeric databases, textual descriptions, GIS maps and layers, photographic images, video files and satellite images. As well, in a broader context, it includes external mathematical models and knowledge bases as input. By having all these input data- and knowledge-bases conveniently accessible, with map-based and/or query-based look-up structures, the exploration of data and model is made easier. For example, data sources may originate from ACCESS, EXCEL, LOTUS, ORACLE, dBASE and so on, and all can be combined and made available as if they were from one large database source (Figure 2).

Tool Assembly Phase

As models and data are assembled, they must be connected to the system backbone, and archived in suitable database and / or program libraries. This task is often a major expenditure of effort, compared to the rest of the work. Data and models are often not as well documented as expected when they are first found. They may exist only in literature form (from the user perspective). Often, simple models are required along with more sophisticated ones, as a "reality check", since they must either agree or the modeler must rationalize the differences.

Hooking up a new model to an existing system can be an onerous programming task. Model inputs require the utmost flexibility of the programmer. The backbone to which the model is connected must have a requisite standard for data input. As well, the data must be available (as needed) to the model. If the modeler finds him/herself stymied by data sharing limitations, then the quality of the whole exercise is compromised.

It is convenient, at this stage, that one can connect to any one of the standard tools or software one is used to handle. However, the problem with large databases, particularly with databases formed from diverse sources, is that each source may require special protocols or at least some specification of the data source to proceed. When too many of these specifications are required, it may become inconvenient. It is again necessary to think of the core structure as offered by a system such as RAISON (Figure 1) in which either the data have been assimilated into one common system, or some standard tools for environmental analysis are made available. For example, as shown in Figure 2, water chemistry data for a watershed from ACCESS (or other sources) can be retrieved by polygon group through a GIS layer of the stream system with a Structured Query Language (SQL) to screen out a special condition for "sample through ice". The data, satisfying the requirements, can be immediately linked to several statistical test tools. These tools are pre-assembled and are ready for the data testing, e.g. frequency analysis, normal and lognormal distributions and various statistical algorithms commonly used in watershed studies. RAISON is also flexible in that if the user wants to link up with external statistical tools at this stage, it can be linked easily through several mechanisms such as the Windows Clipboard.

Trial and Verification Phase

This phase of modeling is tightly connected with tool assembly, and the exercise of trial and verification, on a minute scale, frequently involves the acquisition of new tools. Again, flexibility is required, since a new tool or a new dataset frequently means a significant programming effort.

Flexibility in matching appropriate tools and data sets, particularly for trial cases before production runs of large databases, is desirable in many environmental applications. For example, if one is to establish the regression relationships between different water chemistry variables, one may use some standard statistical package if such relationships can be derived with one or two attempts. However, in many cases, because of differences in sampling methods, analytical chemistry routines and geographical conditions, the relationships can only be derived by matching databases and their meta-files. Visualization of such relationships with respect to site locations on GIS maps or layers will help decide which data sets or parameters will be useful in such analysis.

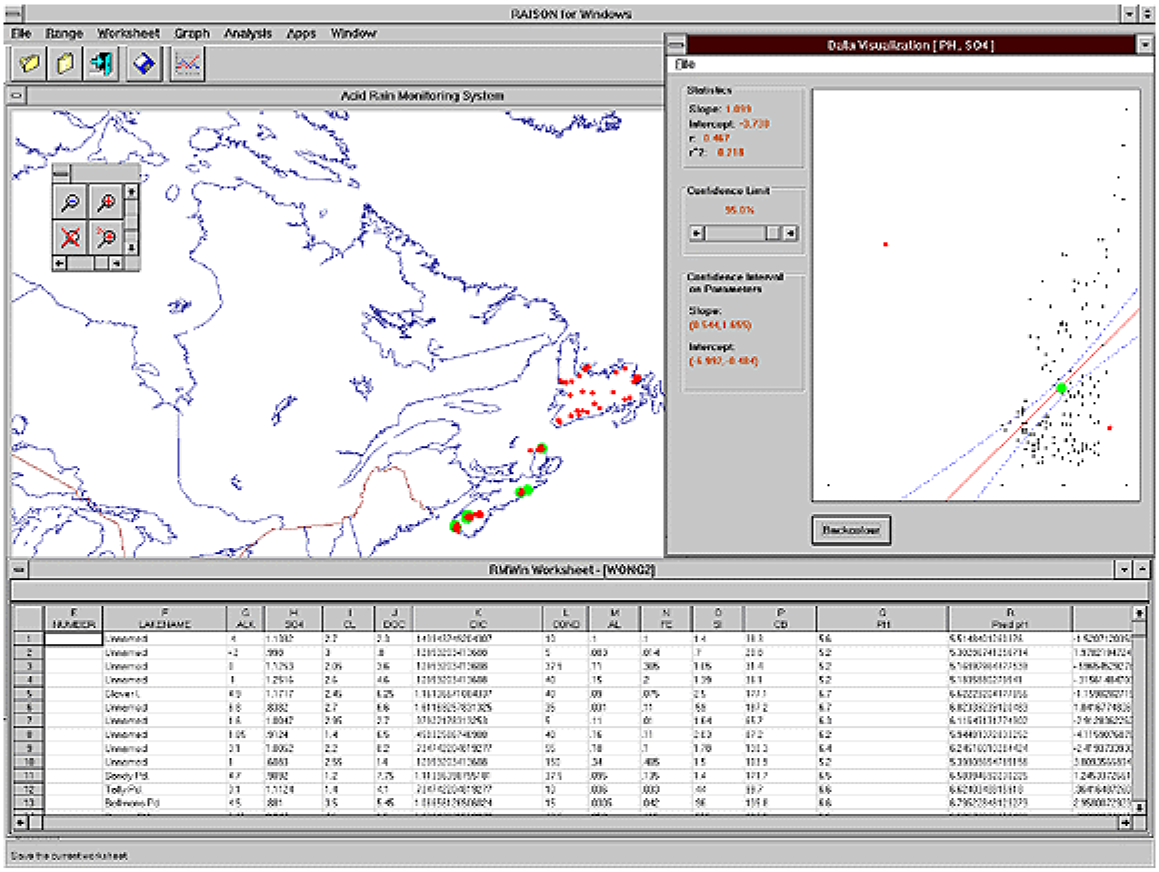


Figure 3. RAISON Support for Trial and Verification

Figure 3 shows an attempt to establish simple regression relationships between, as an example, the water chemistry parameters pH and sulphate for the Atlantic region in Canada. In the RAISON system, the data can be linked or imported into a worksheet (i.e. spreadsheet) in RAISON, from which parameters can be selected. In the graphics module, the two parameters can be plotted as an x-y regression line with the envelopes for the 95% confidence level marked on the scattered plot. By decreasing (or increasing) the confidence level, the envelopes can be widened (or narrowed) to include (or exclude) data points. Keywords or attributes, as discussed for the methods in Figure 2, using meta-files or remark codes on the analytical methods, can be used to screen data points in the trials. For example, by showing the data points inside and outside of the confidence envelope with thematic colouring by sites on a GIS map can help decide quickly which areas or regions are more appropriate for further regression analysis.

Production and Verification Phase

As confidence in the assembly of tools and data improves, rather more effort can be expended on the larger modeling enterprise. Instead of a singular source of pollution, a whole entity (such as a sub-watershed) can be tested. Again, this is inclusive of the previous phase in development, and a return to trial phase or even to assembly phase is not unexpected. Flexibility tends to be required in those earlier phases for measurable progress to occur at this one.

When the confidence level of trial cases reaches a satisfactory limit, production of results can be performed or repeated for different databases. A simple example is to repeat the procedures, as in the case of spreadsheet calculations, for different columns or rows of data. In more complex cases, repetition of processing may need extra programming. For large databases, particularly in the case of different databases, the programming will be very specialized and pertains to individual software or database domains, particularly if data formats are not standardized. This is still very much a new area that requires research and development. In the RAISON system, a special RAISON programming language (RPL) is provided to allow repetitive procedures in production runs.

Report / Presentation Generation Phase

GIS is an important tool for generation of a visual framework for the modeling enterprise. It must be flexible (allowing, for instance a spatial representation of the quality of data coverage or the applicability of models for an assessment of model coverage). To this end, the production of model or data quality assessment is well-handled by contouring or depth-shading. The results provide important feedback to the modeling team. The use of spatial query (such as in Map Mosaic) along with a toolkit for contouring are invaluable at this stage.

Report generation and presentation also require significant import/export facility into word processing and / or presentation tools. It is necessary to use good presentation skills, within the limited framework of an overhead projector or a standard paper page. Sectioning of a large map, generation of tables and diagrams, a history of the work performed are but three examples of necessary tools from the computer software.

Personal computers seem to have the tools for presentation and reporting. They do not necessarily support the high-end GIS. Therefore, once again, data transfer and standards for graphics , etc. must be available and maintainable.

In the RAISON system, the results can be numerical tables, graphical plots, thematic colour maps, or animated sequences of maps. As shown in Figure 1, the output can be textual advice or recommendation, should the expert system tools be used. An example is the case where a complex statistical package, called WATQUAS, developed for the Ontario Ministry of Environment and Energy, is tested in trial cases for provincial water quality monitoring network. During production runs, statistical results were quite voluminous and require further ranking, through the use of expert systems, and were reduced to simple classifications such as high, medium, and low thresholds for policy advisors and the public to understand. Similarly, the aggregation of vast volumes of emission inventory data over the Northern America Continent can be processed with statistical models and summarized for acid rain impact analysis.

Conclusions

On the one hand, the advent of software development has led to easy-to-use linkage technologies for various software including databases. On the other hand, these technologies are not yet sufficient (and will indeed remain so if data continues to grow) to handle large databases, so that a core technology is required that acts as a data filter or transfer points of these databases. The extra programming effort (e.g. map-based or query-based data search) to build this core base must consider not only the databases but also the tools that normally are used to analyze the data. When the core technology can be made generic and applicable for many environmental problems, be they regional or continental in scale, the effort is worth it.

References

Integrating Database, Spreadsheet, GIS, Statistics, Simulation Models in Expert Systems: Experiences with the RAISON System on Micro Computers. NATO, ASI Series, Vol. G 26. Decision Support Systems. Edited by D.P. Loucks and J.R. da Costa. Springer-Verlag Berlin Heidelberg 1991.

Liping Di, R. Suresh, Doug Ilg, and Ted Meyers

EOSDIS Data Models and Example of Implementations

Abstract

As a part of the U.S. Global Change Program, the National Aeronautics and Space Administration (NASA) is developing and implementing its Mission to Planet Earth (MTPE) Program. The foundation of MTPE is the Earth Observation System (EOS). In order to process, distribute, and achieve data from the EOS satellites and to facilitate interdisciplinary research of the Earth system, NASA is developing the EOS Data Information System (EOSDIS) Core System (ECS). EOSDIS is central to the EOS program. It provides the environment in which scientists, educators, policy makers, and other users access and use reliable large-scale data sets of geophysical and biological processes. This paper describes briefly EOSDIS conceptual data models, data format standards, and the implementation of the data models in the Hierarchical Data Format (HDF). The paper discusses in some details about how to implement EOSDIS data models in real data sets by using NASA NSCAT as an example.

Introduction

With the socio-economic development, human economic activities have increasingly affect the global climate and environment. The human-induced global climate and environment changes, such as green-house gas effects, ozone depletion, desertification, deforestation, water and air pollution, have started to limit the development of human society. In recent years, more and more people have realized that we have the responsibility to ensure the planet Earth can sustain the development of human society not only for us but also for our children. In order to reach this goal, we have to understand and predict changes in global climate and environment through studying Earth as a system (Asrar and Dozier, 1994).

As a part of the U.S. Global Change Research Program, the National Aeronautics and Space Administration (NASA) is developing and implementing its Mission to Planet Earth (MTPE) Program. The foundation of MTPE is the Earth Observation System (EOS), which provides the next generation of satellite remote-sensing instruments and platform hardware, a community of funded scientists, and the infrastructure to consolidate data and information from surface campaigns and remote-sensing satellites. The entire EOS program plans to observe the Earth continuously for the next 15 years (Asrar and Greenstone, 1995).

In order to process, distribute, and achieve data from the EOS satellites and to facilitate interdisciplinary research of the Earth system, NASA is developing the EOS Data Information System (EOSDIS) Core System (ECS). EOSDIS is central to the EOS program. It provides the environment in which scientists, educators, policy makers, and other users access and use reliable large-scale data sets of geophysical and biological processes. It also facilitates the interaction and communications among EOSDIS users (Asrar and Dozier, 1995). This paper introduces briefly the EOSDIS conceptual data models, the Hierarchical Data Format (HDF), and the implementation of the conceptual models in HDF. The paper discusses in some details about how to implement the data models in real data sets by using NASA Scatterometer (NSCAT) data as an example.

EOSDIS Conceptual Data Models

Because EOSDIS will process the data from field campaigns, aircraft & ship observations, satellite remote sensing, and the output of scientific models, the data are in various forms. It is unrealistic that EOSDIS deals with the individual data sources individually. Some classification of data forms must be performed so that individual data products can be classified into some limited number of groups of data forms. Thus, software can be developed for the limited number of data groups. In EOSDIS, data are grouped into three generalized models: point, swath, and grid.
Grid: rectangular arrays or data structures for geocoded data (Figure 1).
Swath: simple or complex remote-sensing swath data (Figure 2).
Point: simple or indexed data of geo-located and/or time-tagged point observations or event data (Figure 3).

Each data model is composed of collections of basic data types. Based on the analysis, it is found the data in EOSDIS will be mainly in the following basic data types:
ASCII Text: plain and formatted text for labels, descriptions or documents;
P = V Metadata: "Parameter = Value" for product description and ancillary data;
Tables: standard tables for record-like data; and index tables for indices and variable length counters for ragged arrays;
n-Dim Array: array of records or array of scalars for heterogeneous or homogeneous multi-dimensional data.

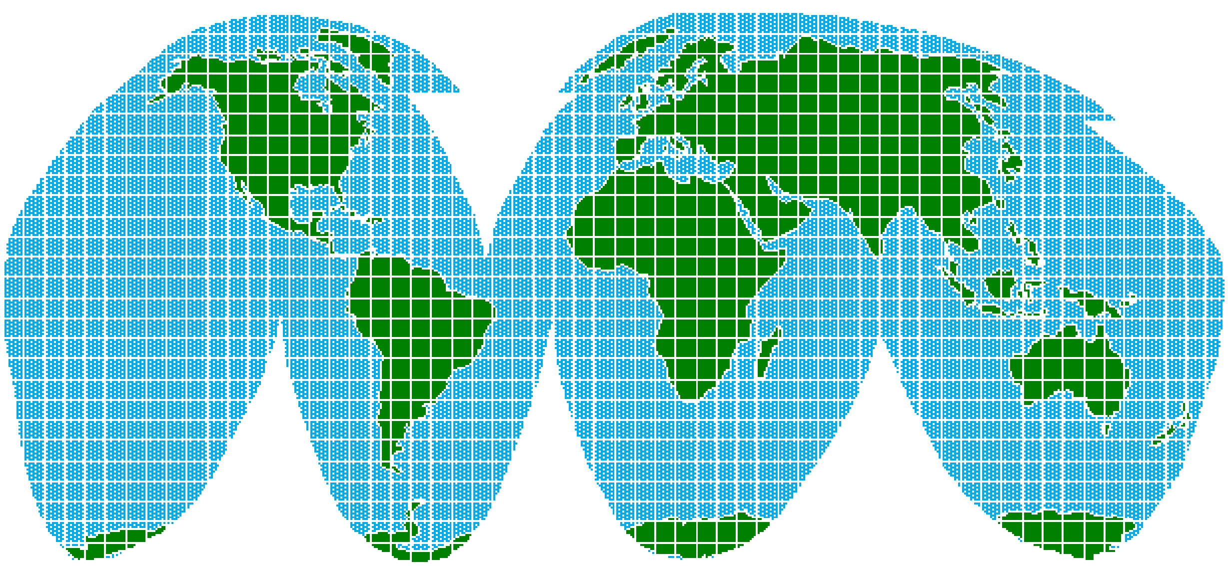


Figure 1. Grid Concept

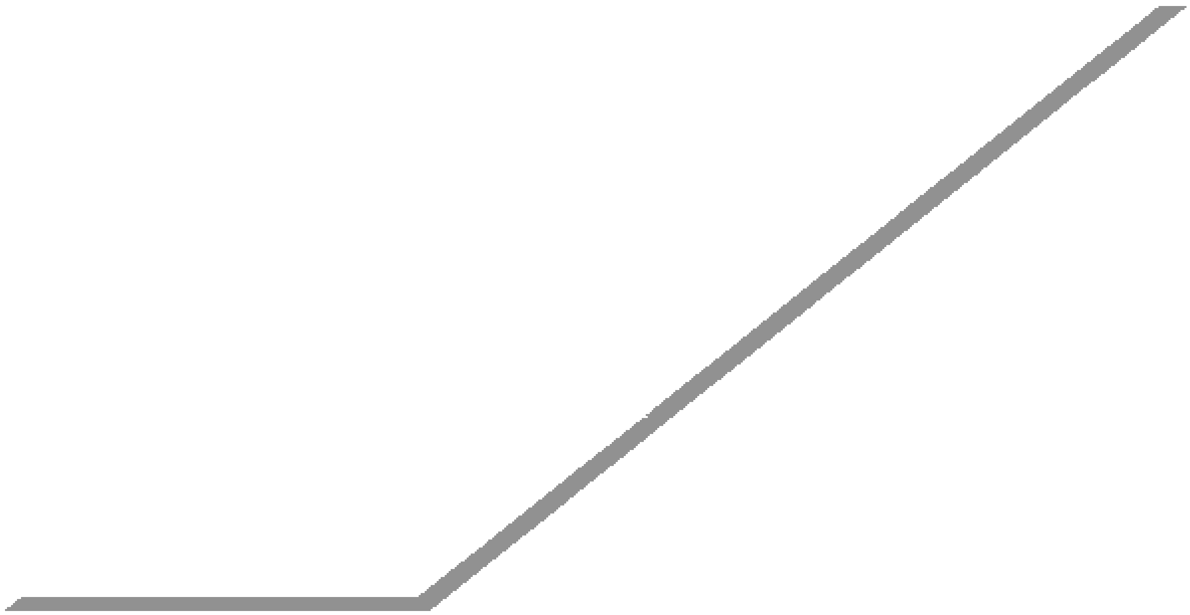


Figure 2. Swath Concept

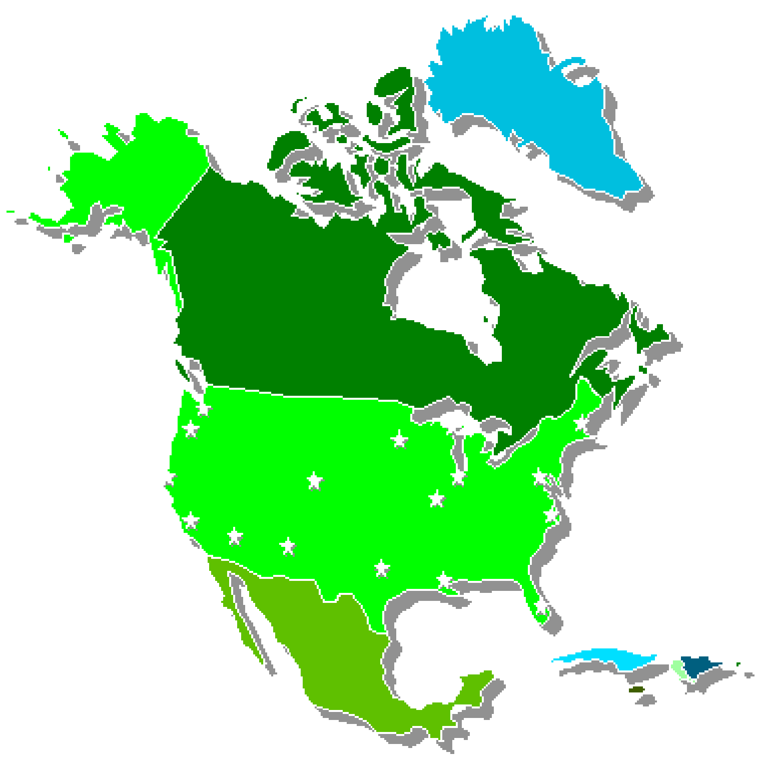


Figure 3. Point Concept

In EOSDIS, data are classified into five levels according to the degree of processing; the higher the level, the higher the degree of the processing (Asrar and Greenstone 1995):
---Level 0-Reconstructured, unprocessed data directly from the instruments at full resolution.
---Level 1A-Reconstructed, unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including geometric and radiometric calibration coefficients and georeferencing parameters.
---Level 1B-Level 1A data that have been processed to sensor units.
---Level 2-Derived geophysical variables at same resolution and location as the Level 1 source data.
---Level 3-Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.
---Level 4-Model output or results from analyses of lower level data.

Among the above data levels, most of data at level 0-2 belong to the swath data model while level 3 and level 4 data belong to the grid data model.

One of the important features of data in EOSDIS is the self-describing. When users obtain data from EOSDIS, the metadata describing the data will also be provided. EOSDIS has defined its core metadata standard (Heller, 1994). The metadata are mostly in text format and stored in "Parameter = Value" form. The data models have to consider for storing the metadata with the instrument data.

Implementation of Data Models in HDF Format

To implement the above conceptual data models in EOSDIS, we have to select one or several physical data formats as the host of the data models. Based on the investigation, it is believed that selecting a single physical format as the EOSDIS standard format will allow Earth science data users to access data more easily and efficiently than multiple formats for performing inter-disciplinary global change research. The advantages of using a single standard format in EOSDIS include:

- promoting science data distribution, exchange, and sharing;
- simplifying data manipulation, analysis, visualization, and management;
- minimizing the cost of system development and maintenance;
- optimizing the system performance;
- supporting long-term preservation of the scientific data.

Because of those advantages, it is ideal that EOSDIS can use a single format as its format standard. However, accommodating those data models in a single data format is not an easy task. The format must be flexible enough to store all possible data types of EOSDIS. Through detailed study, NASA has selected Hierarchical Data Format(HDF) as the standard format (Suresh, 1994, HSTX, 1994). The main reasons of selecting HDF include:

1. Multiple data type support - All data types in EOSDIS can been supported by HDF.
2. Portability - Data in HDF can be ported to many different platforms.
3. Easy to use and implement - HDF provides software library and good documentation.
4. HDF software and documents are free and available on anonymous FTP.
5. Availability of software tools for manipulating and visualizing data in HDF. Most of the tools are free.

6. Compatible with netCDF, a widely used data format in the scientific user community.
7. Supported by many commercial software packages such as IDL.
8. HDF was mainly developed for the scientific user community and has been accepted by scientific user community.
9. Some EOS-related science groups have already selected and recommended HDF.
10. Majority of the DAACs indicated a preference to study HDF

HDF has being developed by the National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign since later 80's for facilitating data access in heterogeneous computing environments (NCSA, 1994). It has six internal data models, including multi-dimensional array (Scientific Data Sets, SDS), table(Vdata), 8-bit raster, 24-bit raster, palette, and annotation (NCSA, 1994). The data in SDS and Vdata can be of any digital types (i.e., float, integer, etc.). Data description (metadata) can be attached to the file and data models. Because HDF allows to store any combinations of the six internal data models in an HDF file, it can meet the requirement of storing the diversity of data in a standard format.

HDF is an object-based data format. Each data object in an HDF file could have a user-defined name. Accompanying with the HDF format, NCSA provides a set of libraries for accessing and managing HDF files. In traditional formats users have to know the exact physical location of an object in the file before the object could be accessed. However, for accessing data in HDF it is not necessary for users to know how a data object is stored in the file and where its physical location is. An application program can directly access a data object through its name by calling HDF library routines. All details of data access are handled by HDF libraries. Therefore, it greatly simplifies the user access to the complicated data structures. The HDF libraries and documents can be obtained freely from ftp://ftp.ncsa.uiuc.edu/.

The process of implementing EOSDIS data models in HDF is to map EOSDIS conceptual models to HDF internal data models and structures. Among HDF data models, the 8-bit raster, 24-bit raster, and palette models are used in EOSDIS mostly in browse and visualization products. The annontation model is somewhat obsolete and is less useful in EOSDIS. Among all data models and objects in HDF, the following strutures are most relevant to the HDF implementation of EOSDIS data models:
---SDS(array):An array of data of any fixed dimensionality (rank) from 1 to 32768 and any one data type. SDSs are currently permitted to have one "unlimited" dimension along which the array can grow indefinitely. The unlimited dimension must be listed as the first dimension in C syntax or the last dimension in FORTRAN syntax.
---Vdata (table):A record based structure wherein record fields may be defined, named, and typed individually. Vdatas are one dimensional arrays of records. Each Vdata may be given an optional name and/or class string for identification purposes.
---Vgroup:A structure for associating sets of data objects. Vgroups may "contain" any HDF objects, including other Vgroups. Like Vdatas, Vgroups may be given an optional name and/or class string for identification purposes.
---Attribute (metadata):A named value or list of values, all the same data type. An attribute can be global (pertaining to the entire file) or local (associated with an individual data object). Currently, only SDSs can have local attributes.

Figures 4, 5, and 6 show the HDF implementation of EOSDIS conceptual data models. In both swath and grid models, instrument data are stored in arrays or tables. The major difference between the two is the implementation of georeferencing. In the swath model, each scan line has a corresponding record in each Vdata in geolocation Vgroup. The record contains information about the georeferencing-related information, such as geolocation of certain pixels within the scan, the start and end time of the scan, orbit parameters, etc. Because of the geocoded nature of data, the geolocation information in the grid model is much simpler than that in the swath model. For each grid, only one geolocation Vdata is required. The Vdata will recorde the map projection related information, which can be used to project grid-cell coordinates back to geographic coordinates. The point data model is implemented by using mainly the Vdata.

Figure 4. Grid model in HDF

Figure 5. Swath model in HDF

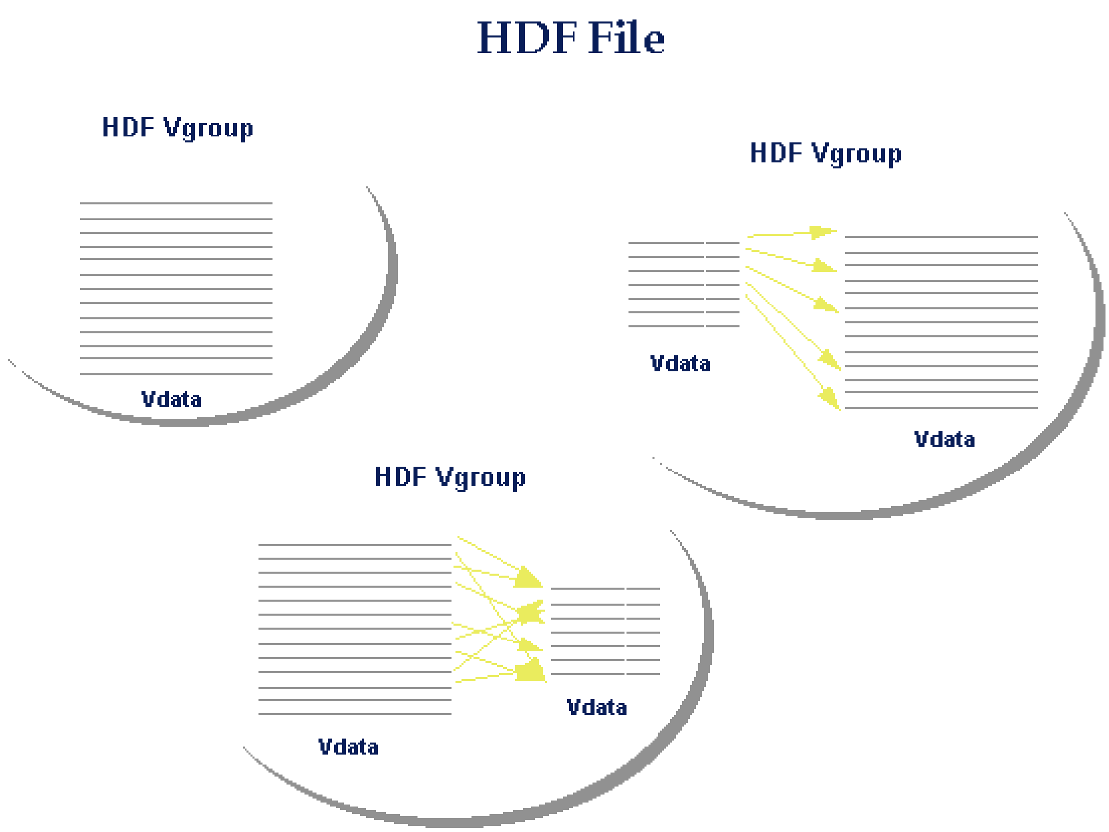


Figure 6. Point model in HDF

Implementation Example and Discussions

The EOSDIS data models have been successfully implemented in data from many EOS data sources such as NSCAT, MODIS, SeaWiFS, and Landsat. In this paper, we discuss the NASA Scatermeter (NSCAT) implementation. NSCAT will be carried by Japanese Advanced Earth Observing Satellite (ADEOS), which will be launched in 1996. The major objective of the mission is to measure the ocean surface wind vectors (Asrar and Greenstone, 1995). The NSCAT data will be archived and distributed by the Distributed Active Archive Data Center (DAAC) at Jet Propulsion Laboratory (JPL). The implementation is performed on the simulated NSCAT data.

Three levels of data, Level 1.7, Level 2, and Level 3, have been implemented in HDF using the EOSDIS swath and grid data models. The Level 1.7 data are ocean-only sigma-0 measurements, grouped by wind vector cell (WVC) rows. The maximum number of data records (WVC rows) is estimated at 820. The nominal number of data records is estimated by taking 70% of the maximum number. The deduction is due to possible land or ice contamination (Di and Benada, 1994 a). The data are in the raw instrument swath geometry and in units of instrument measurement. Because of the swath structure in the data, the swath model has been used to implement the Level 1.7 data in HDF. 26 metadata attributes have been used to describe the data product. Two Vdata (tables) are used to store the record-related metadata; one record in each table for each wind vector row. The instrument data are stored in 21 two, three, or four dimensional arrays (SDS) with each parameter per array. The geolocation information is stored in two arrays and a Vdata. The two arrays are used to store the lat/lon location of each wind vector cell. The Vdata is used to store the index of the each WVC row.

The Level 2 data are wind speed and wind direction in wind vector cells. The number of WVC rows is the same as the Level 1.7 data (Di and Benada, 1994b). Although the data have been coverted to the real physical units of the wind speed and direction, the Level 2 data are still in the swath geometry. Therefore, the swath data model has been used to implement the Level 2 data in HDF. 26 metadata attributes have been used to describe the data product. One Vdata is used to store the record-related metadata. 13 two or three dimensional arrays are used to store the data. The geolocation information are the same as the Level 1.7 data.

The Level 3 data are a set of global averaged maps of wind vector solutions and various secondary variables and statistical descriptors. The averaging interval is one day. Each map contains one averaged parameter. The map grid is in lat/lon coordinate with the spatial resolution of a half degree defined within latitude limits of -75 to 75 degrees and longitude limits of 0 to 360 degrees. The horizontal and vertical map grid dimensions are 720 and 300, respectively. The starting grid row corresponds to the southern-most latitude and the starting grid column corresponds to the western-most longitude within the map region contamination (Di and Benada, 1994c). Because of its grid structure, Level 3 data use the grid data model in the HDF implementation. 24 metadata attributes have been used to describe the data product. 10 two dimensional arrays are used to store the map data with one parameter per array. The geolocation information is stored in a Vdata with the name of the Vdata to be the projection name and the class name of it to be "Geometry". The Vdata includes bin_meth, grid_origin, hsize, max_east, max_north, max_south, max_west, registration, and vsize fields. The hsize and vsize define the grid spatial resolution in horizontal and vertical directions. The max_east, max_north, max_south, max_west define the boundary of the grid spatial coverage. The bin_meth defines the method used for obtaining the grid cell values. The grid_origin defines the relationship between the map coordinate and the grid coordinate. The registration defines the relationship between the grid cell and the geographic location of the cell. The size of each array is defined in the HDF SDS structure as an integrated part of the SDS.

Because of the HDF implementation, NSCAT data can take full advantage of EOSDIS functionality such as data search, browsing, retrieval, and subseting. But from the data user point of views, the simplicity and access speed are also very important for judging if a data format is a good one. Because NSCAT data also exist in record-based native formats besides HDF, we can compare the data access in the native formats with that in HDF to find if the HDF implementation is better than the native one. Through the comparison, we find that data in HDF are much more easy to be retrieved than the native one because in HDF the data are stored as objects. The retrieval speed of data in HDF is comparable with data in the native formats. In the case of retrieving individual parameters, HDF is faster than the native formats because it does not require to read in whole file for unpacking before the needed data can be obtained.

Because of large amount of data from satellite, the storage efficiency of a data format is also very important. In order to pack different types of data into a record and keep records in constant length in whole file, redundancy and waste space are common in record-orientated native formats. The HDF is based on data object models which treats a data set as a collection of individual data objects and has no concept of data records as such. Therefore, HDF implementation of NSCAT data can remove any redundancy and waste space existing in the native formats. Although an HDF file includes its own tags, reference ids and other object-related information in the file, HDF files use less storage space than native NSCAT files. Table 1 lists the sizes of three simulated NSCAT native data files and their corresponding HDF files. The table shows that the HDF files are consistently smaller than the native files.

Data Level	Level 1.7	Level 2	Level 3
Size in Native(Bytes)	7455576	704704	4324320
Size in HDF(Bytes)	6851994	648698	3911367

Table 1. The storage size for NSCAT data in HDF and in native formats.

The implemention of remotely-sensed data in HDF by using EOSDIS data models normally takes two steps; mapping conceptually the individual data components in a data set into HDF data objects and writing the software to form the HDF data file. Based on our experiences, the most time consumptive step is the conceptual mapping. The mapping is also very important because it affects the performance of data access in HDF files. Once the mapping is done, the software development for writing the data in HDF is relatively easy.

References

Asrar, G., and Dozier, J., 1994. EOS: Science Strategy for the Earth Observing System. American Institute of Physics Press, Woodbury, New York 11797.

Asrar, G., and Greenstone, R., (1995). 1995 MTPE EOS Reference Handbook. NASA Goddard Space Flight Center, Greenbelt, MD 20771.

Di, L. and Benada, R., (1994a). NASA Scatterometer (NSCAT) Project Science Data Processing System, Level 1.7 Data Software Interface Specification (SIS-2) - HDF Version. JPL D-12059. Jet Propulsion Laboratory, CA.

Di, L. and Benada, R., (1994b). NASA Scatterometer (NSCAT) Project Science Data Processing System, Level 2 Data Software Interface Specification (SIS-2) - HDF Version. JPL D-12060. Jet Propulsion Laboratory, CA.

Di, L. and Benada, R., (1994c). NASA Scatterometer (NSCAT) Project Science Data Processing System, Level 3 Data Software Interface Specification (SIS-2) - HDF Version. JPL D-12061. Jet Propulsion Laboratory, CA.

Heller, D., 1994. Proposed ECS Core Metadata Standard, Release 2.0. Hughes Applied Information System, Landover, MD.

HSTX, 1994. EOSDIS Information Management System Users Manual. Hughes STX Corporation, Greenbelt, MD.

NCSA, 1994. HDF Reference Manual. University of Illinois at Urbana- Champaign.

Suresh, R., 1994. EOSDIS Version 0 Data Product Implementation Guidelines. Hughes STX Corporation, Greenbelt, MD.

Liping Di , R. Suresh, and Doug Ilg
Hughes STX
7701 Greenbelt Road, Suite 400
Greenbelt, MD 20770
E-mail: lpd@ulabsgi.gsfc.nasa.gov
Phone: 301-441-4104, FAX: 301-441-1853

Ted Meyers
NASA Goddard Space Flight Center (NASA/GSFC), Code 505, Greenbelt, MD 20771.

William K. Michener and Paula F. Houhoulis

Identification and Assessment of Natural Disturbances in Forested Ecosystems: The Role of GIS and Remote Sensing

ABSTRACT

Ecologists, natural resource managers, and environmental modelers require accurate identification and assessment of the ecological impacts associated with natural disturbances in order to better understand forest ecosystem patterns and processes, to mitigate adverse impacts, and to support modeling activities. At the landscape scale, natural disturbances include fire, insect infestation, hurricanes and tropical storms, flooding, and high winds. Because many natural disturbances result in highly variable damage over large areas, accurate assessment of impacts may be difficult and time-consuming.

In this review, the role (including relevant examples) of GIS and remote sensing in identification, assessment, and monitoring of natural disturbances is discussed. Data sources, image and GIS processing techniques (data transformations, change detection algorithms), and accuracy assessment are reviewed. A case study is presented whereby four change detection approaches are evaluated for their effectiveness in discriminating vegetation changes associated with flooding in a forested ecosystem. Finally, recommendations for natural disturbance assessment are identified.

INTRODUCTION

Natural disturbances play an important role in regulating forest ecosystem structure and function, as well as affecting diverse plant and animal populations and communities. Because vegetation typically exhibits abrupt changes in physiognomy and spectral characteristics in response to acute disturbances, environmental scientists are increasingly using digital images obtained by satellite remote sensors that can detect these changes over broad spatial, temporal, and spectral scales. Change detection analyses, employing Geographic Information System (GIS) coverages and satellite data obtained prior to and following a disturbance, have been used to assess vegetation responses to drought (Peters *et al.* 1993, Jacobberger-Jellison 1994), insect outbreaks (Muchoney and Haack 1994), dust storms (Chavez and MacKinnon 1994), high winds (Cablk *et al.* 1994, Johnson 1994), deforestation (Foody and Curran 1994), and other disturbances.

Various data sources and analytical approaches differing in mathematical complexity, processing and analysis intensity, and classification technique have been used to detect vegetation change. Many studies have relied upon less computationally intensive post-classification change detection techniques using images or GIS coverages from one or two dates (Aldrich 1975, Sirois and Ahern 1989, Gardner *et al.* 1991, Cablk *et al.* 1994, Dobson *et al.* 1995, Olsson 1995). Recently, principal components analysis (PCA), various vegetation indices, and logic rules have been implemented utilizing multitemporal satellite data (Bauer *et al.* 1994, Muchoney and Haack 1994, Jensen *et al.* 1995, Walsh and Townsend 1995).

Objectives of this paper are to examine how GIS and remote sensing have been used to assess ecological impacts of natural disturbances in forested ecosystems, and to review data sources, relevant GIS and image processing techniques, and accuracy assessment procedures. A case study is presented whereby four change detection approaches are evaluated for their ability to discriminate vegetation changes associated with flooding of a forested ecosystem. Finally, we recommend approaches that appear promising for future change detection studies and identify several research challenges.

USE OF GIS AND REMOTE SENSING FOR IDENTIFYING, ASSESSING, AND MONITORING NATURAL DISTURBANCES IN FORESTED ECOSYSTEMS

Natural Disturbances in Forested Ecosystems

Forested ecosystems are constantly undergoing change. Many of the changes (e.g., succession, responses to climate change, etc.) are directional and occur incrementally over long periods of time. Other changes (e.g., treefall, deforestation, etc.) are more acute, ranging in size from small gaps to entire forests, and act to reset areas back to earlier successional states or entirely alter the ecosystem state altogether.

Many types of disturbances and forest change occur at broad spatial scales and are of special interest to environmental scientists and resource managers. Frequently, because of the size of the areas affected, forest disturbances and their impacts are assessed using remotely sensed data, GIS data, and appropriate change detection approaches. For example, change detection analyses employing remotely sensed data have been used to assess forest impacts of high winds (Cablk *et al.* 1994, Johnson 1994), fire (Lopez-Garcia and Caselles 1991), salinization (Cablk *et al.* 1994), climate change (Awaya *et al.* 1994), and deforestation and harvesting (Tucker *et al.* 1984, Sader 1987, Bauer *et al.* 1994, Foody and Curran 1994).

Insect infestations are frequently monitored using remotely sensed data since they can cause widespread forest defoliation and significantly affect commercial interests. Examples include: mountain pine beetle, *Dendroctonous ponderosae* Hopk. (Sirois and Ahern 1989); hemlock looper, *Lambdina fiscellaria fiscellaria* (Franklin 1989); spruce budworm, *Choristoneura fumiferana* Clemens (Buchheim *et al.* 1984); gypsy moth, *Lymantria dispar* L. (Rohde and Moore 1974, Nelson 1983, Ciesla *et al.* 1989, Muchoney and Haack 1994); and pear thrips, *Taeniothrips inconsequens* Uzel (Vogelmann and Rock 1989).

Data Sources

Sources of data for forest change detection studies vary in spectral, spatial, and temporal resolution. When available, high resolution color or color infrared aerial photographs are used to detect large-scale (local) changes or, more frequently, for assessing accuracy of small-scale (regional) changes identified from lower spatial resolution satellite data. The effectiveness of satellite data for detecting different types of forest change depends to a large extent upon the spatial resolution of the satellite sensor, which can range from 10 m (SPOT panchromatic) to 1 km (NOAA Advanced Very High Resolution Radiometer, AVHRR). For example, SPOT multispectral data (20 m resolution) have been used to identify relatively small forest stands that have been defoliated by insects (Muchoney and Haack 1994), Landsat Thematic Mapper (TM) data (30 m resolution) have been used to assess moderate forest damage resulting from the high winds and saltwater storm surge associated with Hurricane Hugo (Cablk *et al.* 1994), and AVHRR (1 km resolution) data have been used to assess deforestation occurring at regional to continental scales (Tucker *et al.* 1984). Choice of satellite data for a particular change detection study must also be based on availability, cost, and spectral resolution.

Image and GIS Processing Techniques

Various analytical approaches differing in complexity, computational intensity, and ease of interpretation have been employed in change detection studies. Although there is increasing interest in fuzzy logic (e.g., Gong 1993) and other new methods, most forest change detection studies have employed a relatively small number of techniques. Some of the most common change detection approaches include: (1) post-classification change detection differencing; (2) spectral-temporal change classification; (3) data transformations (e.g., Normalized Difference Vegetation Index, NDVI); (4) principal components analysis (PCA); (5) image differencing; and (6) change vector analysis (CVA).

Post-classification change detection differencing involves comparing classes from two or more digital data sets on a pixel or area (polygon) basis. Prior to change detection, each digital image must be independently classified by the analyst using supervised or unsupervised approaches. Examples of post-classification change detection are provided by Wickware and Howarth (1981), Estes *et al.* (1982), and Muchoney and Haack (1994).

Spectral-temporal change classification is based on classification of a single merged data set containing spectral data from multiple dates. When the images are obtained immediately prior to and after a disturbance or on anniversary dates (similar phenology, etc.), then only those areas that have undergone change will be significantly different and the remainder of the data will be similar. An example of spectral-temporal change classification is provided by Muchoney and Haack (1994).

Although broad-scale land use changes may often be readily detected using raw spectral data, more subtle changes such as vegetation stress may be more difficult to identify. In such cases, specific band ratios or band combinations may facilitate change detection. For example, Cablk *et al.* (1994) employed image differencing of Landsat TM 7/4 band ratios and NDVI data ((IR-R)/(IR+R)) to accurately identify forest stands affected by high winds and saltwater intrusion. Many of the most widely applied vegetation indices are reviewed by Perry and Lautenschlager (1984) and Nilsson (1995).

Principal components analysis (PCA) is a multivariate statistical technique that isolates inter-image change by transforming linear combinations of band data into components that account for the maximum (1st component) and successively lower proportions (2nd and higher order components) of variance among image layers. Applications of PCA techniques are reviewed by LeDrew (1987) and Jiaju (1988).

Image differencing is based on band-by-band subtraction of digital numbers (DNs) using two images (e.g., band1(yr1)-band1(yr2), etc.). Frequently, a median value is added to the differenced data set to eliminate negative values, prior to standard unsupervised classification. Examples of classifications derived by image differencing are provided by Robinove *et al.* (1981), Cablk *et al.* (1994), and Muchoney and Haack (1994).

Change vector analysis is an empirical method used to detect radiometric changes based on multidate satellite data, and is characterized by vectors representing the magnitude and direction of changes present in the data (Malila 1980, Michalek *et al.* 1993). Malila (1980) successfully applied CVA to brightness and greenness bands derived from Landsat TM data to detect changes in forest extent due to harvesting and regrowth. Other applications of CVA are described in articles by Michalek *et al.* (1993), Johnson (1994), and Lambin and Strahler (1994).

Accuracy Assessment

Classifications derived from remotely sensed images are subject to error and uncertainty. In classifying an image, the spectral response of a pixel, representing a fixed area on the ground defined by the resolution of the sensor, is used to assign it to one of a number of classes using various classification techniques. To assess classification accuracy, reference (ground truth) data are needed for a number of sample locations for each class. Accuracy is defined in terms of misclassifications, where a pixel is assigned to the wrong class. Misclassifications are usually presented in the form of a matrix which is referred to as a confusion or error matrix (Table 1). The error matrix can be used to generate various statistics that characterize the accuracy of a classification technique. For example, the overall accuracy compares the number of pixels correctly classified (those appearing on the diagonal of the matrix) to the total number of pixels sampled (see Table 1). However, this statistic can be misleading since a certain number of correctly classified pixels are expected to occur by chance alone. The Cohen's Kappa or Khat statistic allows for chance, and ranges from 0 in the case of the most confused classification to 1 in the case of the most accurate classification (Table 1). Other statistics that can be generated from the error matrix include errors of omission (producer's error) and errors of commission (user's error). These are based on individual classes, dividing the number of pixels that are incorrectly classified by either the column or row totals, respectively (Table 2). Additional discussion of accuracy assessment techniques can be found in articles by Congalton (1988, 1991), Czaplewski (1994), and Goodchild (1994).

TABLE 1. Example error matrix depicting observed classes versus actual classes of forest types.

Forest Type	Pine	Hardwood	Scrub	Row Total
Pine	**43**	7	3	53
Hardwood	7	**14**	9	30
Scrub	0	8	**12**	20
Column Total	50	29	24	

Overall Accuracy = 69 / 103 = 0.699

Khat = (103(69) - 4000) / ((103)2 - 4000) = 0.470*

$$^{*}K_{hat} = (N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+} \bullet x_{+i}) / (N^2 - \sum_{i=1}^{r} x_{i+} \bullet x_{+i})$$

Where r is the number of rows in the matrix, x_{ii} is the number of observations in row i and column i, x_{i+} and x_{+i} are the marginal totals of i and column i respectively, and N is the total number of observations (Bishop *et al.* 1975).

TABLE 2. Errors of omission (producer's error) and commission (user's error) using Table 1 as a reference.

Forest Type	Omission Error	Commission Error
Pine	(7+0) / 50 * 100 = 14%	(7+3) / 53 * 100 = 19%
Hardwood	(7+8) / 29 * 100 = 52%	(7+9) / 30 * 100 = 53%
Scrub	(3+9) / 24 * 100 = 50%	(0+8) / 20 * 100 = 40%

Various change detection techniques may be applied in any one study, resulting in multiple classifications. Goodchild *et al.* (1992) proposed a general error model, called a Probability Vector Model (PVM), for obtaining estimates of uncertainty in land cover maps. In their model, each classification scheme is treated as a "realization" and combined to form a data layer from which the uncertainty associated with a class at any point (or pixel, *ij*) is represented by a vector of probabilities {p*ij*1, p*ij*2,..., p*ij*n} defining the probability that a pixel belongs to each class 1 through *n* (Goodchild *et al.* 1992, Goodchild 1994).

CASE STUDY: DETECTING VEGETATION RESPONSES TO FLOODING IN A FORESTED ECOSYSTEM

Tropical Storm Alberto presented an opportunity to examine the utility of satellite data for assessing ground cover vegetation responses to flooding in a natural forested ecosystem. Minimal wind and storm surge damage accompanied Alberto as it made landfall on the Florida panhandle. However, due to weak steering currents, the storm remained relatively stationary over southwestern Georgia and southeastern Alabama for a period of six days (July 2-7, 1994). Rainfall was especially heavy (up to 53 cm) in the Flint and Ocmulgee River basins in southwestern Georgia and flood discharges on tributaries and mainstems of the two rivers exceeded 100-year flood discharges along most stream reaches (Stamey 1995). Natural habitats in the two basins are characterized by longleaf pine trees (*Pinus palustris*) and wiregrass (*Aristida stricta*), the dominant ground cover species.

Although satellite data have been used to reconstruct regional flood history (Nagarajan *et al.* 1993), map water boundaries and changes in major wetland habitat types (Wickware and Howarth 1981, Walsh and Townsend 1995), and relate agricultural crop damage to severity of flooding (Yamagata and Akiyama 1988, Yamagata *et al.* 1988), none have related ground cover vegetation responses to flooding in natural terrestrial ecosystems. The purpose of this case study was to evaluate four different change detection approaches for their ability to discriminate vegetation responses to differential severity of flooding. Relatively sparse canopy coverage in longleaf pine stands, typical of many forest types that occur in xeric habitats, enabled satellite sensors to detect spectral characteristics of the dense ground cover vegetation. Extensive ground surveys supported evaluation of the effectiveness of different change detection approaches, and factors that affect their accuracy.

Study Area

Ichauway is a 115 km2 ecological reserve that is located in Baker County in southwest Georgia, 45 km southwest of Albany (Figure 1). The site is located along the Flint River at its confluence with Ichawaynochaway Creek. Approximately 22 km of Ichawaynochaway Creek and 19 km of the Flint River, a brownwater stream originating in the Georgia Piedmont region, are located within the reserve.

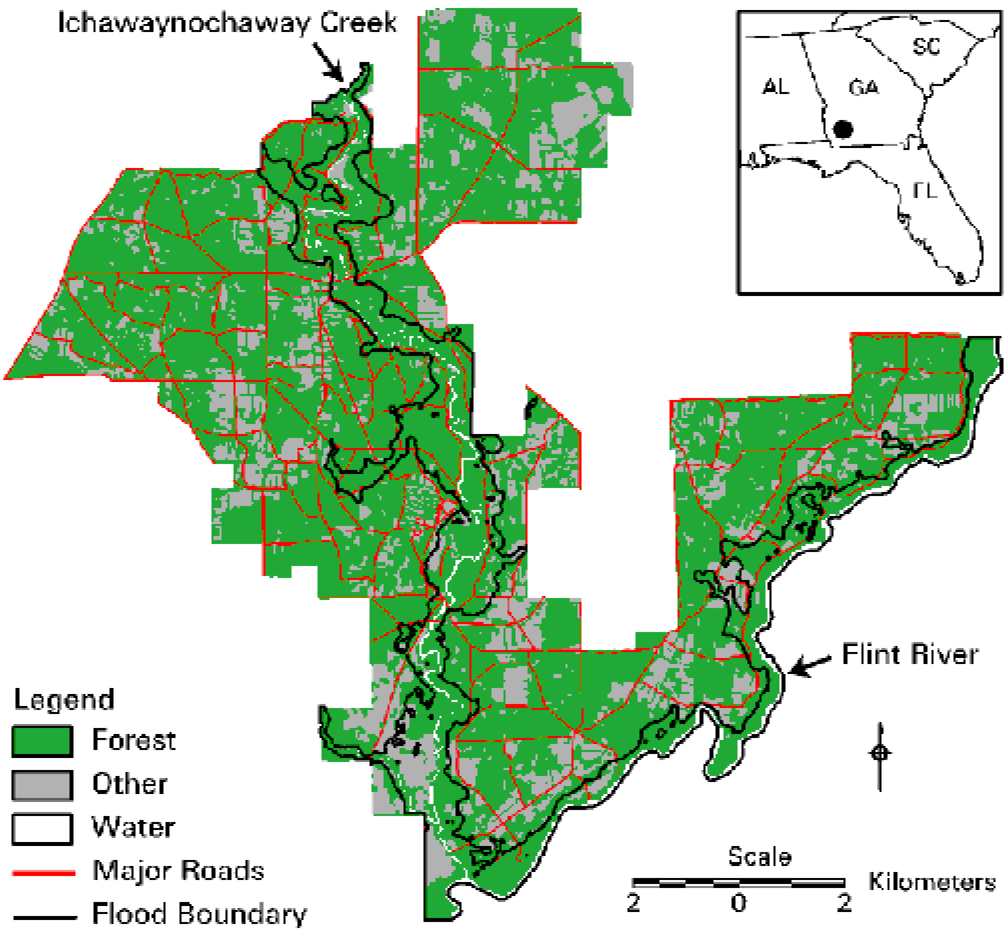


FIGURE 1. Location of Ichauway study site showing generalized land cover and extent of flooding associated with Tropical Storm Alberto.

Remotely Sensed Data

Multispectral (XS) images of the study area were acquired from SPOT Image Corporation for September 28, 1994 and October 5, 1990. All images were processed to level 1B, were predominantly cloud-free, and had incidence angles less than 7.5 degrees. A linear regression model was used to correct for relative differences in atmospheric conditions between the two image dates (Jensen *et al.* 1995). Digital numbers were sampled using a 3x3 window, in 3 types of areas; dark water bodies, dense conifer forest stands, and bright bare soils. Because their appearance was consistent from year to year, the same areas in each image were used as sample sites. Mean values from the sample areas were regressed using the 1990 image as 'master' and the 1994 image as a 'slave'. Regression equations were applied to the 'slave' image using the ERDAS Imagine Modeler (Table 3).

TABLE 3. Regression Equation Used to Normalize Radiometric Characteristics of the 1994 Data with the 5 October 1990 SPOT XS Data.

Date	Band 1 (Green)	Band 2 (Red)	Band 3 (Near IR)
Sept 28, 1994	y = 0.93(x)+ 2.76, r2 = 0.98	y = 1.02(x)+ 0.29, r2 = 0.99	y = 0.98(x)- 1.40, r2 = 0.99

Twenty three ground control points (GCPs) digitized from 7.5 minute USGS topographic quadrangles (USGS quads) were used to rectify the October 5, 1990 SPOT-XS image to a Universal Transverse Mercator (UTM) map projection (RMSE = 0.29 pixels / 5.91 m). The 'slave' image was similarly rectified using GCPs obtained from the 1990 rectified image (1990 RMSE = 0.33 pixels / 6.63 m). Images were resampled to a 20 m pixel size using a nearest neighbor resampling technique (Jensen *et al.* 1993). To insure that the data layers used in this analysis were co-registered, the relative error between the images and the ancillary data layers was estimated by taking the difference between ten well distributed checkpoints (road intersections) whose coordinates were recorded from the rectified images and the GIS transportation layer (Wolter *et al.* 1995). The relative error (RMSE) was less than 5 m in each case.

Ancillary Data Layers

Ancillary data layers were used to assess classification accuracy and derive other data layers such as image masks that were used in the change detection analyses. Three ancillary data layers used in this study (landcover, groundcover, and transportation routes) were interpreted from 1:12,000 scale color infrared (CIR) aerial photographic transparencies. Data were transferred using a vertical sketchmaster to USGS quads, digitized, and attributed using ARC/INFO. Landcover classification attributes included detailed descriptions of species composition, age class, and stand density for all forested areas. Groundcover attributes included primary and secondary cover types and vegetation density. The transportation layer included linear features such as fire-breaks, state and county maintained roads, and highways. To insure that the layers were co-registered, the road network and water bodies were used as a coincident line layer, keeping these features consistent in each of the other photo-interpreted layers. Elevation spot heights and 5 ft contours were digitized from USGS quads. In a few cases, 5 ft contours were interpolated from 10 ft contours using GRASS software.

The progression of the floodwaters was monitored on site during July 1994, and maximum water levels were recorded at approximately 350 locations along Ichawaynochaway Creek and the Flint River. High water levels were surveyed with Trimble Global Positioning System (GPS) Pro XL and Basic Plus receivers and differentially corrected ($\pm$ 2 m) to a known Community Base Station. Maximum water levels were used to derive a flood boundary map by overlaying the points on topography, and extrapolating along contour lines between the points to form a polygon.

The ground cover and land cover layers were combined, and used to define a mask containing only forested areas, excluding agriculture, urban areas, roads, water, non-forested wetlands, and regenerating forest stands. This process reduced the potential confusion between flood-damaged vegetation and other land uses and changes (e.g., crop rotations).

***In Situ* Reference Data**

One hundred and twelve sites (approximately 650 m2 per site) dominated by wiregrass groundcover were surveyed throughout the flooded area to quantify vegetation damage. Each site contained three randomly chosen plots where groundcover mortality was assessed using a 1 m quadrate divided into a 10 X 10 grid at 10 cm intervals. Presence of bare ground, detritus, and wiregrass condition (dead, live, or recovering) were recorded at all points and the data from the three plots were averaged and converted into percentages. Each site was surveyed using the GPS techniques described earlier.

Image Classification

Four change detection techniques were evaluated in this study. The first, spectral-temporal change classification (S-TCC), is based on unsupervised classification of the spectral data for the two dates. The second approach, S-PCA, was applied to a six-band merged spectral data set consisting of spectral data from both dates. Third, image differencing (S-ID) of the spectral bands from the two images was performed prior to unsupervised classification. Fourth, image differencing was based on differences in NDVI values (NDVI-ID) observed prior to and following the flood.

S-TCC was based on unsupervised classification of a single multidate data set that contained the six bands from the two dates. ISODATA, an algorithm available in Imagine 8.1, was used in the unsupervised classification. Fifty unsupervised signatures were extracted. The number of iterations was adjusted as necessary to achieve a 0.96 convergence level. The large number of classes provided relatively narrow clusters, that were visually inspected and re-classed as flooded or non-flooded based on the spatial distribution of each class in relation to the areal extent of flooding.

S-PCA was based on a merged six band data set containing all spectral bands from the two images. Examination of the eigenstructure of the transformed data indicated that the first four components accounted for almost 99% of the spectral variability among the images (Table 4). Components five and six were attributable to atmospheric and sensor variations. Therefore, only the first four components were retained for classification.

TABLE 4. Eigenstructure for Multitemporal PCs Based on Spectral Data.

Band		1	2	3	4	5	6
1990	1	0.16	0.47	-0.20	-0.21	-0.79	0.19
	2	0.15	0.67	-0.20	-0.41	0.55	-0.16
	3	0.70	-0.32	-0.62	0.12	0.07	-0.02
1994	1	0.07	0.25	0.05	0.49	-0.17	-0.81
	2	0.05	0.40	-0.00	0.72	0.19	0.53
	3	0.67	0.02	0.73	-0.10	0.00	0.05
Eigenvalues		46.71	34.57	12.64	5.31	0.86	0.35
% Variance		46.51	34.42	12.58	5.29	0.86	0.34
% Cum. Var.		46.51	80.93	93.51	98.80	99.66	100.00

Two variations of the image differencing approach were evaluated in this study. First, image differencing (S-ID) of the spectral bands from the two images was applied prior to unsupervised classification (as described for S-TCC). In the second approach, NDVI-ID, 1994 NDVI values were subtracted from the 1990 NDVI values to obtain a value for each pixel that represented a magnitude and direction of change. In the resulting image data set, values that are negative or close to zero indicate areas where greenness increased in 1994 or remained relatively unchanged, whereas positive values represent areas exhibiting a decrease in greenness in 1994. For this case study, positive values greater than 11 DNs were classed as flooded.

Accuracy Assessment

Data from the 112 ground survey sites were used in the accuracy assessment. GPS-derived coordinates for each of the ground survey sites were given unique identification numbers (id) and imported into ARC/INFO as a point coverage. Flood classes generated from the different change detection methods (binary masks where 1 = flood) and the flood zone layer were imported into GRID from Imagine 8.1. Site survey data (point coverage) were rasterized using a 1 m cell size and combined with each of the flood class methods and flood zones using the GRID statement gpsflood = COMBINE (gpspoints, floodzone, method(1)...method(n)). The raster value attributes (VAT) contained in the combined layer, 'gpsflood', were related and transferred to the point attribute table (PAT) of the site survey layer based on the GPS-id. The PAT was then output to an ASCII text file to be used for statistical analysis.

The ASCII file was input into the Statistical Analysis System (SAS) and merged with the groundcover mortality data. The sites were classified as live or dead based on the percentage groundcover dead (40% = dead). The number of dead sites classified as flooded versus those incorrectly classified as non-flooded, and the number of live sites classified as non-flooded versus those incorrectly classified as flooded, were used as a measure of how well each of the change detection methods performed. Overall accuracy and Kappa Coefficients (Khat) were calculated for each method using techniques described by Congalton *et al.* (1983) and Congalton (1991).

A Probability Vector Model (PVM) was used to facilitate visualization of accuracy assessment for the four change detection schemes employed. The four binary images were combined to derive probability values that ranged from 0 to 1, representing the proportion of times a pixel was classified as flooded by the different methods or realizations. Thus, a probability value equal to 0.75 indicates that the pixel was classified as flooded by three of the four techniques evaluated.

Results

Overall accuracy and Kappa Coefficient statistics (Khat) were used to compare the different change detection techniques (Table 5). Overall accuracy was high for all techniques and ranged from 0.607 to 0.750. However, Khat values exhibited significant variation, ranging from 0.266 to 0.487. Both measures indicated that spectral-temporal change classification (also known as layered temporal change classification) was least effective in discriminating flood-affected vegetation (Table 5a; Figure 2a).

TABLE 5. Accuracy assessment of four change detection techniques used to detect vegetation responses to flooding.

	No. Dead Sites		No. Live Sites			
Method	Det.	Und.	Det.	Und.	Accuracy	Khat
a. S-TCC	40	6	28	38	0.607	0.266
b. S-PCA	9	7	31	35	0.625	0.291
c. ID	34	12	45	21	0.705	0.409
d. NDVI-ID	33	13	51	15	0.750	0.487

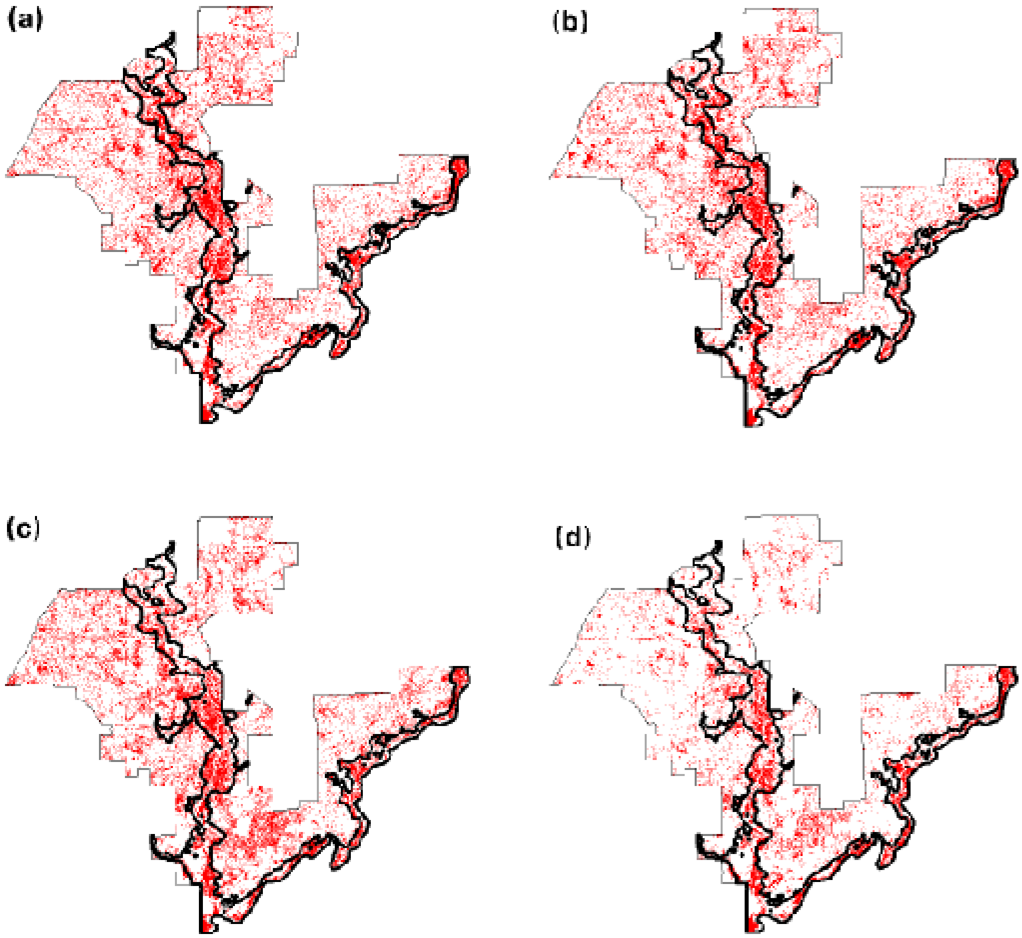


FIGURE 2. Classification maps of Ichauway study area showing results of change detection analyses based on: (a) spectral-temporal change classification (S-TCC); (b) principal components analysis of all spectral bands (S-PCA); (c) image differencing of spectral bands (S-ID); and (d) image differencing of NDVI data using a user defined threshold (NDVI-ID).

Classification accuracy and Khat values exhibited marginal improvement when classification was based on S-PCA (Table 5b; Figure 2b). Image differencing of spectral data (S-ID) resulted in a marked improvement in classification accuracy over the first two methods (Table 5c; Figure 2c). However, image differencing based on NDVI (NDVI-ID) was the most effective technique for discriminating vegetation responses to flooding; accuracy was highest and Khat exhibited a two-fold increase over S-PCA (Table 5d; Figure 2d). Closer examination of the NDVI data indicated that digital numbers (DNs) were similar in non-flooded areas in 1990 and 1994, but were substantially lower (> -10 DNs) in the flooded area in 1994, in comparison to 1990 (Table 6). These findings indicate that ground cover vegetation exhibits a marked spectral response to flooding which is best exemplified as a decrease in NDVI in affected areas.

TABLE 6. Mean NDVI Values for the 1990 and 1994 SPOT XS Data.

Image Date	Entire Site	Non-flooded	Flooded	Standard

				Deviation
10/05/90	190.2	189.7	192.5	1.9
09/28/94	186.4	187.6	181.7	4.1

A map based on PVM was used to visually assess sources of confusion in the classification process and to better understand the factors that affect our ability to discriminate flood-affected vegetation (Figure 3). Many of the flood-affected vegetation zones along the Flint River and Ichawaynochaway Creek that were identified by all four techniques occurred in areas that experienced highest current velocities and deepest waters as well as in localized depressions where standing water remained for several days following the flood. Within the flood boundary, 82% of the area that appeared to be affected was classified as "flooded" by at least one or more of the four techniques that are depicted in Figure 3. Furthermore, almost half (49%) of the "flooded" area was discriminated by at least two of the four techniques and 33% of the "flooded" area was detected by all four approaches. Thus, it is apparent that there is considerable agreement among the techniques in identifying vegetation changes within the flood zone. In contrast, most (20%) of the area outside the flood boundary that was misclassified as "flooded" was discriminated by only one of the four techniques. Only 6% of the area outside the flood boundary that was misclassified as "flooded" was identified by all four techniques; with many of the larger clusters (shown in red; Figure 3) frequently being associated with vegetated wetlands that were affected by the excessive precipitation.

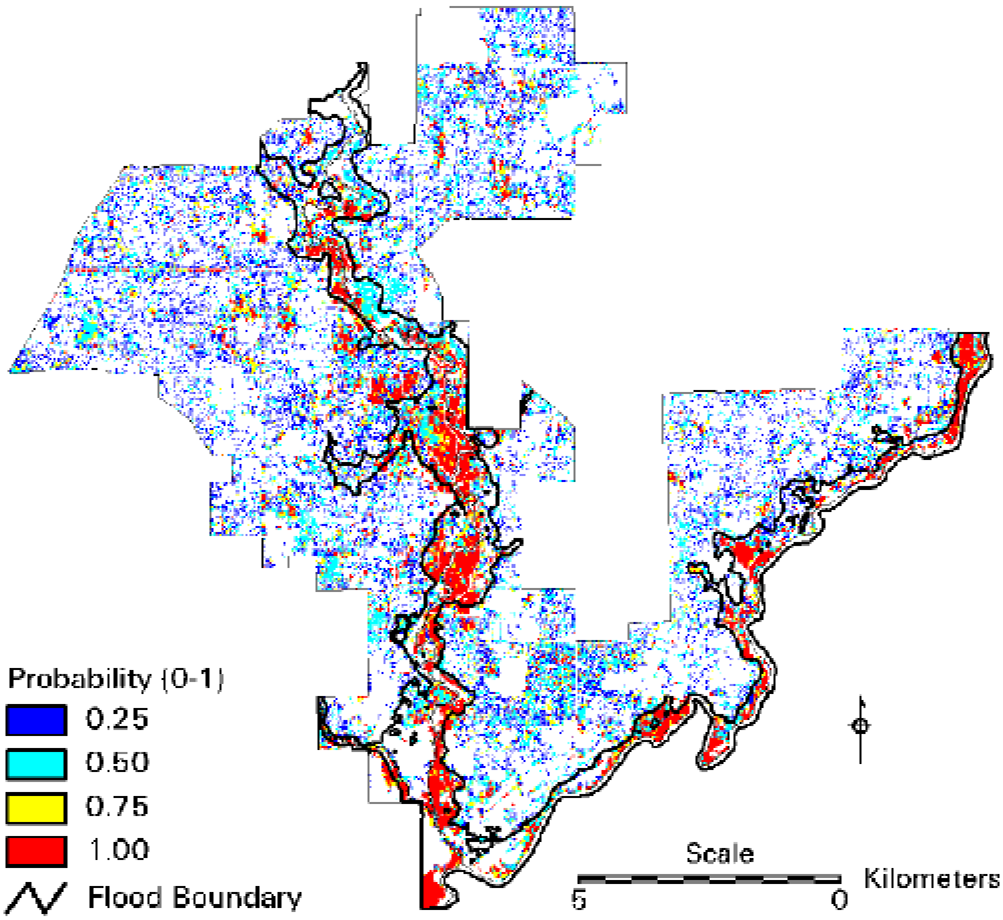


FIGURE 3. Map based on probability vector model showing proportion of times that a pixel was classified as flooded by four different change detection techniques.

Discussion

Several factors might be expected to constrain flood impact assessments. First, the timing of the satellite data coverage relative to river conditions frequently leads to underestimates of the severity and areal extent of flood inundation (Blasco *et al.* 1992). Second, dense vegetation canopy and the complex relationship between hydrologic and phenologic cycles may confound vegetation spectral responses within the floodplain (Walsh and Townsend 1995). Third, accuracy assessments generally require reliable post-flood ground truth data, adequate digital elevation models, or other data that are often lacking or inadequate. Finally, vegetation may exhibit a lagged response to secondary flood-related factors (anaerobiasis, waterlogging, etc.) in areas that do not directly experience the most intense erosion and scouring.

In this study, cloud-free satellite data could not be acquired until almost three months after the flood. Relatively sparse longleaf pine overstory coupled with a dense ground cover community dominated by a single species (wiregrass) facilitated change detection analyses. Overall classification accuracy exceeded 60 per cent for the unsupervised methods used in this study, suggesting that all techniques were effective for discriminating vegetation responses to flooding. However, the approximate two-fold range in Kappa Coefficient Statistics (Khat = 0.266 - 0.487) indicated that some methods outperformed others when chance agreement was removed. Specifically, it is possible to generalize that: (1) a classification based on changes occurring in all spectral bands, S-TCC, was least effective in discriminating vegetation changes related to the flood; (2) S-PCA offered marginal improvement in classification accuracy over temporal change classification; and (3) image differencing represented the most effective method for discriminating flood-affected vegetation.

CONCLUSION AND RECOMMENDATIONS

Change detection studies frequently focus on very abrupt changes related to alterations in land use (e.g., deforestation) and broad-scale natural disturbances (e.g., hurricanes, drought, insect outbreaks). The case study presented above demonstrates the feasibility for using satellite data to detect and monitor relatively subtle responses in vegetation dynamics to natural disturbances in forested ecosystems. This capability offers significant potential for increasing our understanding of ecosystem- and landscape-scale responses to natural disturbances as well as assessing changes in vegetation dynamics related to climatic variability and global climate change.

The effectiveness of change detection studies of natural disturbances may be significantly affected by temporal, spatial, and spectral resolution of the data, as well as the availability of ancillary (relevant GIS coverages) and ground truth data. When change detection analysis is based on data acquired immediately prior to and following a discrete disturbance event, spectral change may be related to ecological changes with a reasonably high degree of certainty. Otherwise, spectral changes associated with a specific disturbance may be confounded with land use change, annual phenological differences, climate, and other factors that differ between the pre- and post-disturbance imagery. Identification of changes occurring over long periods of time (e.g., succession, climate change, etc.) may require time series of images and greater automation of change detection analyses. Automation of change detection, however, requires that baseline conditions be defined prior to assessment of change. Unfortunately, very little is known about what constitutes "normal" conditions for an area. For example, can a single satellite image be considered to represent normal conditions or would the range of spectral variability present during wet and dry years (two or more images) serve more effectively as a baseline for assessing future change?

Spatial resolution may be of less concern when major disturbances (e.g., hurricanes, droughts, etc.) or land use changes (e.g., deforestation of tropical rain forests) occur over extremely large areas. More frequently, however, forest responses to disturbances are highly variable and occur in patches. Thus, the ability to discriminate local areas of change will be related to patch sizes and sensor resolution (Townshend 1981). For example, despite the 20 m sensor resolution, Sirois and Ahern (1989) found that the smallest areas affected by Mountain Pine Beetles that could be detected with SPOT XS data ranged from 1 to 2 ha in size, and contained trees with 80 to 100 per cent damaged crowns. Unfortunately, little is known about the relationships among sensor spatial and spectral resolution, minimal patch size detected, and type, variability and magnitude of damage in natural forest ecosystems.

The effectiveness of alternative change detection approaches for assessing forest disturbances is rarely evaluated within the context of a single study. Although approaches that work well within a single comparative study may not necessarily apply in other ecosystem types or for other types of disturbance, results of such studies may provide guidance or, at least, a starting point for other disturbance assessments. Muchoney and Haack (1994) evaluated four change detection approaches using multitemporal SPOT High Resolution Visible (HRV) data, ranging from standard post-classification change differencing to more analytically complex image differencing and PCA techniques, for identifying hardwood forest defoliation caused by gypsy moth infestation. In their study, overall accuracy ranged from 0.61 (post-classification, spectral-temporal) and 0.63 (PCA) to 0.69 (image differencing of spectral bands). In the case study presented in this paper, image differencing was similarly the most effective of the techniques evaluated. One interesting result, however, was the improvement in classification accuracy when image differencing was based on NDVI data as opposed to the spectral bands for the two images. Of the various techniques evaluated by Muchoney and Haack (1994) and in the case study presented in this paper, image differencing represents a relatively straightforward technique that could be easily automated for specific areas of interest.

Accuracy assessments within the context of a study and inclusion of relevant statistics (especially Kappa Coefficients) in subsequent publications can support evaluation of the effectiveness of change detection approaches for specific applications, and may facilitate future research efforts. Maps derived from probability vector modeling can be used to visually interpret classifications resulting from multiple change detection methods, and may also lead to a better understanding of factors that affect accuracy. For example, large areas that are similarly classified by all or most methods can be easily identified. Areas of disagreement (lower probability scores) may indicate mixed pixels or class uncertainty. Consistent areas of disagreement may, for example, delineate changes occurring in a particular land cover class (e.g., vegetation senescence in bottomland hardwood habitats) that are being confused with the change of interest (e.g., insect defoliation of conifer stands). Although such sources of confusion may be easily visualized, interpreted, and corrected, they may not be apparent using standard classification accuracy assessment techniques. Goodchild *et al.* (1992) further discuss how this modeling approach can be used to obtain standard errors associated with area estimates.

Once baseline conditions are established and change is detected in a "new" satellite data coverage, ground or aerial photography-based assessments are essential for determining whether "significant" spectral change is ecologically meaningful. In some cases (e.g., deforestation, severe defoliation associated with wind, insect outbreaks, etc.), spectral change may be easily and directly related to vegetation change, regardless of the change detection approach employed. In other cases, the ability of different change detection approaches to discriminate vegetation changes may be affected by forest stand characteristics, land cover, soil characteristics, and so forth. Logistic multiple regression represents a powerful analytical technique that may prove useful for evaluating different change detection methods and designing ground verification studies. Logistic regression has frequently been used to investigate the relationship between response probabilities of binary and ordinal response variables, and the explanatory variables (Hosmer and Lemeshow 1989). Binary response variables (e.g., defoliated, unaffected) and ordinal response variables (e.g., no effect, moderate defoliation, severe defoliation) arise in many studies of ecological disturbances. Logistic regression analysis may be effective for investigating the relationship between the spectrally defined response probability and the potential explanatory variables (e.g., degree of canopy closure, stand condition, soil moisture class, etc.). Results of such analyses could be used to reduce the number of attributes that are monitored in the field, thereby reducing sampling costs.

ACKNOWLEDGMENTS

We thank the numerous students and technicians who participated in ground survey efforts, especially Derek Fussell; and Frank Miller, Mary Grace Chambers and Linda Garnett of the Mississippi Remote Sensing Center, and Jean Brock (JWJERC) for GIS support. This project was funded by the National Science Foundation (DEB-9520878).

REFERENCES

Aldrich, R.C. (1975) Detecting disturbances in a forest environment, *Photogrammetric Engineering & Remote Sensing* 41: 39-48.

Awaya, Y., T. Hame, and N. Tanaka (1994) Vegetation change detection using Landsat data in northern Finland. *Proceedings for the National Institute of Polar Research Symposium* 7: 270-282.

Bauer, M.E., T.E. Burk, A.R. Ek, P.R. Coppin, S.D. Lime, T.A. Walsh, D.K. Walters, W. Befort, and D.F. Heinzen (1994) Satellite inventory of Minnesota forest resources. *Photogrammetric Engineering & Remote Sensing* 60(3): 287-298.

Bishop, Y., S. Fienberg, and P. Holland (1975) *Discrete Multivariate Analysis-Theory and Practice*. Cambridge, Massachusetts: MIT Press.

Blasco, F., M.F. Bellan, and M.U. Chaudhury (1992) Estimating the extent of floods in Bangladesh using SPOT data. *Remote Sensing of Environment* 39: 167-178.

Buchheim, M.P., A.L. Maclean, and T.M. Lillesand (1984) Forest cover type mapping and spruce budworm defoliation using simulated SPOT imagery. *Proceedings of the 1984 SPOT Symposium*, pp. 259-266.

Cablk, M.E., B. Kjerfve, W.K. Michener, and J.R. Jensen (1994) Impacts of Hurricane Hugo on a coastal forest: Assessment using Landsat TM data. *Geocarto International* 2: 15-24.

Chavez, P.S., and D.J. MacKinnon (1994) Automatic detection of vegetation changes in the southwestern U.S. using remotely sensed images. *Photogrammetric Engineering & Remote Sensing* 60(5): 571-583.

Ciesla, W.M., C.W. Dull, and R.E. Acciavatti (1989) Interpretation of SPOT-1 color composites for mapping defoliation of hardwood forests by gypsy moth. *Photogrammetric Engineering and Remote Sensing* 55(10): 1465-1470.

Congalton, R.G. (1988) A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing* 54(5): 593-600.

Congalton, R.G. (1991) A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37: 35-46.

Congalton, R.G., R.G. Oderwald, and R.A. Mead (1983) Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering & Remote Sensing* 49(12): 1671-1678.

Czaplewski, R.L. (1994) Variance Approximations for Assessments of Classification Accuracy. Research Paper RM-316, Fort Collins, Colorado: United States Department of Agriculture, Forest Service.

Dobson, E.L., J.R. Jensen, R.B. Lacy, and F.G. Smith (1995) A land cover characterization methodology for large area inventories. *ACSM/ASPRS Proceedings*, pp. 786-795.

Estes, J.E., D.Stow, and J.R. Jensen (1982) Monitoring land use and land cover changes. Pages 100-110 *in* C. Johannsen and J. Sanders, eds. *Remote Sensing for Resource Management*. Ankeny, Iowa: Soil Conservation Society of America.

Franklin, S.E. (1989) Classification of Hemlock Looper defoliation using SPOT HRV imagery. *Canadian Journal of Remote Sensing* 15(3): 178-182.

Foody, G.M., and P.J. Curran (1994) Estimation of tropical forest extent and regenerative stage using remotely sensed data. *Journal of Biogeography* 21: 223-244.

Gardner, L.R., W.K. Michener, B. Kjerfve, and D.A. Karinshak (1991) The geomorphic effects of Hurricane Hugo on an undeveloped coastal landscape at North Inlet, South Carolina. *Journal of Coastal Research* 8: 181186.

Goodchild, M.F. (1994) Integrating GIS and remote sensing for vegetation analysis and modeling: Methodological issues. *Journal of Vegetation Science* 5: 615-626.

Goodchild, M.F., G.Q. Sun, and S. Yang (1992) Development and test of an error model for categorical data. *International Journal Geographic Information Systems* 6: 87-104.

Gong, P. (1993) Change detection using principal component analysis and fuzzy set theory. *Canadian Journal of Remote Sensing* 19(1): 22-29.

Hosmer, D.W., and S. Lemeshow (1989) *Applied Logistic Regression*. New York, New York: John Wiley and Sons, Inc.

Jacobberger-Jellison, P.A. (1994) Detection of post-drought environmental conditions in the Tombouctou region. *International Journal of Remote Sensing* 15(16): 3138-3197.

Jensen, J.R., D. Cowen, J.D. Althausen, S. Narumalani, and O. Weatherbee (1993) An evaluation of the CoastWatch change detection protocol in South Carolina. *Photogrammetric*

Engineering & Remote Sensing 59(6): 1039-1046.

Jensen, J.R., K. Rutchey, M.S. Koch, and S. Narumalani (1995) Inland wetland change detection in the everglades water conservation area 2A using a time series of normalized remotely sensed data. *Photogrammetric Engineering & Remote Sensing* 61(2): 199-209.

Jiaju, L. (1988) Development of principal component analysis applied to multitemporal Landsat TM data. *International Journal of Remote Sensing* 9(12): 1895-1907.

Johnson, R.D. (1994) Change vector analysis for disaster assessment: A case study of hurricane Andrew. *Geocarto International* 1: 41-45.

Lambin, E.F., and A.H. Strahler (1994) Indicators of land-cover change for change-vector analysis in multitemporal space at coarse spatial scales. *International Journal of Remote Sensing* 15(10): 2099-2119.

LeDrew, E. (1987) Application of principal components analysis to change detection. *Photogrammetric Engineering and Remote Sensing* 53(12): 1649-1658.

Lopez-Garcia, M.J. and V. Caselles (1991) Mapping burns and natural reforestation using Thematic Mapper data. *Geocarto International* 1: 31-37.

Malila, W.A. (1980) Change vector analysis: an approach for detecting forest changes with Landsat. *1980 Machine Processing of Remotely Sensed Data Symposium*, pp. 326-336.

Michalek, J.L., T.W. Wagner, J.J. Luczkovich, and R.W. Stoffle (1993) Multispectral change vector analysis for monitoring coastal marine environments. *Photogrammetric Engineering & Remote Sensing* 59(3): 381-384.

Muchoney, D.M., and B.N. Haack (1994) Change detection for monitoring forest defoliation. *Photogrammetric Engineering & Remote Sensing* 60(10): 1243-1251.

Nagarajan, R., and G.T. Marathe (1993) Identification of flood prone regions of Rapti River using temporal remotely-sensed data. *International Journal of Remote Sensing* 14(7): 1297-1303.

Nelson, R.F. (1983) Detecting forest canopy change due to insect activity using Landsat MSS. *Photogrammetric Engineering & Remote Sensing* 49: 1303-1314.

Nilsson, H. (1995) Remote sensing and image analysis in plant pathology. *Canadian Journal of Plant Pathology* 17: 154-166.

Olsson, H. (1995) Reflectance calibration of TM data for forest change detection. *International Journal of Remote Sensing* 16(1): 81-96.

Perry, C.R., and L.F. Lautenschlager (1984) Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment* 14: 169-182.

Peters, A.J., B.C. Reed, M.D. Eve, and K.M. Havstad (1993) Satellite assessment of drought impact on native plant communities of Southeastern New Mexico, U.S.A. *Journal of Arid Environments* 24: 305-319.

Robinove, C.J., P.S. Chavez, D. Gehring, and R. Holmgren (1981) Arid land monitoring using Landsat albedo difference images. *Remote Sensing of Environment* 11: 133-156.

Rohde, W.G., and H.J. Moore (1974) Forest defoliation assessment with satellite imagery. *Proceedings of the Symposium on Remote Sensing of Environment* 2(9): 1089-1104.

Sader, S.A. (1987) Digital image classification approach for estimating forest clearing and regrowth rates and trends. *Proceedings of IGARSS '87 Symposium* pp. 209-213.

Sirois, J., and F.J. Ahern (1989) An investigation of SPOT HRV data for detecting recent mountain pine beetle mortality. *Canadian Journal of Remote Sensing* 14(2): 104-108.

Stamey, T.C. (1995) Floods in Central and Southwestern Georgia in July 1994. *Proceedings of the 1995 Georgia Water Resources Conference*, pp. 313-316.

Townshend, J.R.G. (1981) The spatial resolving power of earth resources satellites. *Progress in Physical Geography* 5: 32-55.

Tucker, C.J., B.N. Holben, and T.E. Goff (1984) Intensive forest clearing in Rondonia, Brazil, as detected by satellite remote sensing. *Remote Sensing of Environment* 15: 255-261.

Vogelmann, J.E., and B.N. Rock (1989) Use of Thematic Mapper data for the detection of forest damage caused by the pear thrips. *Remote Sensing of Environment* 30: 217-225.

Walsh, S.J., and P.A. Townsend (1995) Comparison of change detection approaches for assessing a riverine flood hydroperiod. *ACSM/ASPRS Proceedings*, pp. 134-143.

Wickware, G.M., and P.J. Howarth (1981) Change detection in the Peace-Athabasca Delta using digital Landsat data. *Remote Sensing of Environment* 11: 9-25.

Wolter, P.T., D.J. Mladenoff, G.E. Host, and T.R. Crow (1995) Improved forest classification in the northern lake states using multi-temporal Landsat imagery. *Photogrammetric Engineering & Remote Sensing* 61(9): 1129-1143.

Yamagata, Y., and T. Akiyama (1988) Flood damage analysis using multitemporal Landsat TM data. *International Journal of Remote Sensing* 9(3): 503-514.

Yamagata, Y., C. Wiegand, T. Akiyama, and M. Shibayama (1988) Water turbidity and perpendicular vegetation indices for paddy rice flood damage analysis. *Remote Sensing of Environment* 26: 241-251.

William K. Michener, Ph.D.

Associate Scientist, Landscape Ecology, wmichene@jonesctr.org and

Paula F. Houhoulis, M.A.

Remote Sensing Analyst, Landscape Ecology, phouhoul@jonesctr.org

---

Joseph W. Jones Ecological Research Center,

Route 2, Box 2324, Newton, GA 31770

ph#:(912) 734-4706, fax#:(912) 734-6650

Janet Franklin and John Stephenson

Integrating GIS and remote sensing to produce regional vegetation databases: attributes related to environmental modeling

---

Abstract

We are using image processing of satellite imagery, and ecological gradient models implemented as cartographic models in a GIS, to produce regional (totaling about 2 million hectares) fine-scale (2-ha minimum mapping unit) digital vegetation maps for the US Forest Service in Southern California. Attributes mapped include vegetation community type, and forest cover and structure (canopy size class). Automated image segmentation is used to delineate mapping units corresponding to vegetation stands from satellite imagery. The databases will be used for a number of natural resources and land management planning applications. We illustrate the usefulness and limitations of these data for evaluating the effect of forest fragmentation on the viability of the California Spotted owl (*Strix occidentalis occidentalis*), a Forest Service management indicator species. A habitat suitability model for the Spotted Owl was developed and applied to the vegetation database, and evaluated for its ability to capture known owl nesting sites in the San Bernardino Mountains. The resulting habitat quality map will be used as input to a spatially explicit simulation model of owl population dynamics, which includes habitat-specific demographic parameters.

Introduction

The management of ecosystems requires inventory and monitoring of large areas of natural landscapes at fine scales. Increasingly, modeling is being used as a research and management tool to examine spatio-temporal processes such as land use conversion, natural disturbance, resource harvesting, and species dynamics (hence, this conference). A lot has been written about trade-offs among spatial, temporal, and categorical resolution in geographic databases, often with reference to existing and planned satellite sensors, issues of data storage and processing, issues of scale and accuracy, and so forth. A number of continental- and global-scale land cover, vegetation, and biodiversity databases are being developed or planned, for example continental and global land cover maps (Loveland et al. 1991, Townshend et al. 1991), MODIS land cover products (Running et al. 1994), global vegetation databases used in modeling the general circulation of the atmosphere (Henderson-Sellers 1994), Landsat Pathfinder datasets (Maiden 1994, James and Kalluri 1994), and the Gap Analysis Program biodiversity datasets (Scott et al. 1993). However, these data may lack the spatial detail and categorical attributes required for some ecosystem management applications.

In this paper we will discuss a vegetation database developed for a large area (2 million ha) at a relatively fine spatial scale (2 ha MMU), with moderate taxonomic detail (vegetation types mapped at the series level) but including information on forest cover and canopy structure. Mapping is based on image processing of Landsat satellite data, and predictive mapping based on terrain and other variables in a GIS. The database is expensive to produce, but we will demonstrate that data of at least this categorical and spatial resolution are required over large regions (several million ha) for ecosystem management. Ecosystem processes operate across large spatio-temporal scales, therefore modeling is essential to addressing landscape management questions (Sample 1994, J.F. Franklin 1993, 1995). The example we will use, from our ongoing research, is that of a habitat suitability model developed for the California Spotted Owl and applied to the vegetation database. The resulting habitat quality map will be used as input to a spatially explicit simulation model of owl population dynamics, which includes habitat-specific demographic parameters.

Certain salient aspects of the mapping methods will be reviewed briefly, as they relate to the integration of remote sensing with GIS, and issues of data structures. Then, the data requirements of a spatially explicit wildlife population simulation model will be outlined, to demonstrate the necessity for stand-specific estimates of vegetation structure and composition. Finally, the preliminary results from our attempt to parameterize and implement a habitat suitability model for the California Spotted Owl will be presented. The habitat model is evaluated based on its ability to predict known owl nesting site locations in the San Bernardino Mountains. Errors can result from lack of precision and accuracy in the vegetation database. However, a database lacking the attributes or spatial resolution of ours would not be useful at all for this type of modeling.

Stand-based Mapping of Vegetation Type and Structure Using Remote Sensing and GIS

As growing numbers of researchers and resource managers rely on digital geographic data, and look to remotely sensed imagery as a source of data for their GIS, issues of geographic data models and remote sensing scene models (Goodchild 1992, 1994, Strahler et al. 1986), as well as image processing algorithm development, are central to research on the integration of remote sensing and GIS. Our mapping methods integrate advanced image processing algorithms (segmentation, canopy modeling, mixture modeling), applied to Landsat Thematic Mapper (TM) imagery, with simple cartographic modeling using GIS, to produce fine-scale digital vegetation maps for large areas. These methods have been described in detail elsewhere (Woodcock et al. 1994a, Franklin and Woodcock, submitted, and references therein, and see Strahler 1981, and Franklin et al. 1986 for background).

Briefly, our methods result in a digital vegetation map with the following attributes:

- a minimum mapping unit corresponding to a vegetation stand derived automatically from satellite multispectral data using image segmentation;
- a vegetation life form label (conifer, hardwood, brush, grass) assigned to each stand based on multispectral image classification;
- a vegetation series label assigned using simple gradient models for each life form developed from field data and applied to topographic variables derived from digital elevation models;
- a canopy cover estimate for the primary forest type derived from a forest canopy reflectance model applied to the TM data;
- a cover estimate for the secondary forest type derived from spectral mixture modeling; and
- canopy structure (crown size class) estimated for each forest stand using the canopy model, or interpreted from air photos.

Importantly, each attribute listed above is interactively edited by an image analyst based on air photo interpretation, usually resulting in the correction of 5- 20% of the labels for each attribute.

Central to our mapping methods, attributes are derived from fine-grained raster data (e.g., 30-m TM and DEM pixels), but the fundamental mapping unit is the "stand" composed of contiguous pixels, derived automatically using image segmentation, and with a minimum size specified (Woodcock and Harward 1992). White and Running (1994, 690) noted that errors arise at all steps of topographic and satellite data processing, and "these errors can be reduced, assuming they are random, by aggregating the land units which have similar environmental characteristics." Averaging our estimates of categorical and continuous attributes over areas larger than the single grid cell tends to reduce errors in the estimates of those spatially autocorrelated geographic phenomena.

The accuracy assessment for the digital vegetation map referred to in this paper has not been completed. However, the accuracy of other forest maps based on the same methods have been evaluated using an innovative approach based on fuzzy set theory (Gopal and Woodcock 1994), resulting in overall accuracy for the life form attribute of about 95%, and for the vegetation series and cover labels of about 60-80% with the greatest errors occurring among similar classes (Woodcock et al., 1994b). The database was costly to produce, very roughly $0.4/ha, not including the initial purchase of satellite imagery, digital elevation models, digital line graphs, and other digital coverages (streams and water bodies, roads, land ownership and administrative boundaries). Photointerpretation-based map editing and field data collection represent the greatest costs.

Data Requirements for Spatially Explicit Population Modeling

Spatially explicit population models (SEPMs) are simulation models that provide a means to assess how habitat patterns influence population dynamics and viability (Dunning et al. 1995). Species' populations are frequently distributed in disjunct patches of suitable habitat embedded in a large landscape matrix. Spatial and temporal variation in habitat patches (e.g. size, structure, vegetation type) and underlying landscape patterns (e.g. distance between patches, dispersal corridors) can profoundly influence population viability (Soule et al. 1988). Habitat patterns influence populations because demographic rates (i.e. survivorship, fecundity, and dispersal success) are affected by variations in habitat quality and configuration (Pulliam 1988). High quality habitats generally have positive demographic rates and contribute to population growth (i.e. "source" habitats). Conversely, low quality habitats can have negative rates and promote population decline (i.e. "sinks") (Pulliam 1988).

SEPMs address the influence of habitat patterns by incorporating digital maps of the landscape (that can be changed over time to simulate dynamic processes) and habitat-specific demographic parameters into a temporal population model. This makes it possible to investigate how changes in landscape configuration, brought on by human land use practices and natural disturbance and successional processes, will affect wildlife populations over time. It has been noted, however, that problems associated with parameter estimation and model complexity are magnified for spatially explicit models, which may give the "illusion of exactitude in the absence of hard information" (Doak and Mills 1994). This is particularly true for the estimation of habitat-specific survivorship rates and dispersal behavior, since there are very few species for which we have empirical data on these parameters. Nevertheless, a SEPM designed to describe the spatial dynamics of wildlife populations in fragmented habitats with a minimum of life history data could at least be used to separate species for whom habitat pattern alone is a good predictor of carrying capacity from those whose population dynamics require a more detailed understanding of mobility and demographics (Schumaker, unpublished manuscript). To date most SEPMs have been developed for relatively well-studied endangered species (Pulliam et al. 1992, McKelvey et al. 1993).

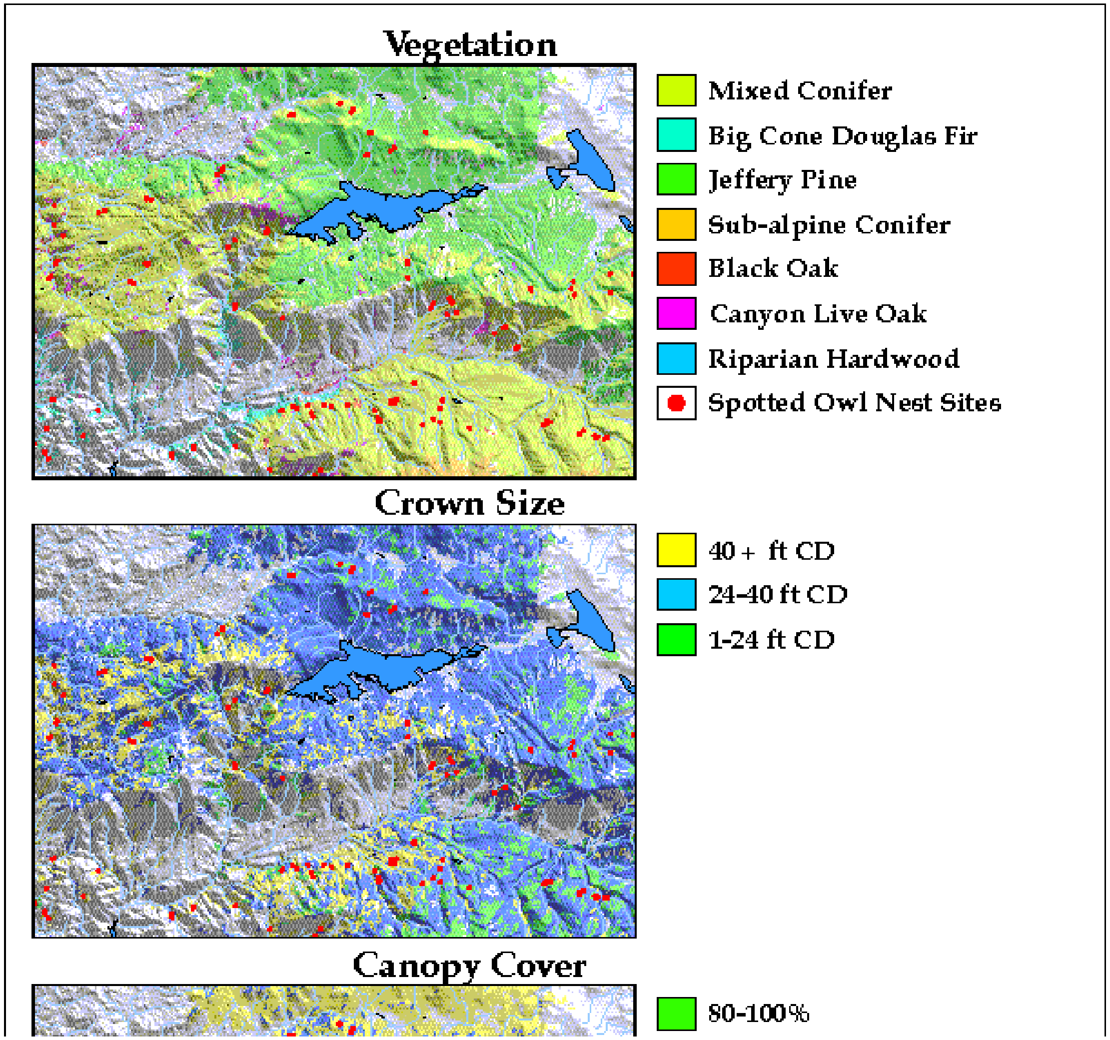
One objective of our ongoing research is to evaluate the effect of present and projected habitat fragmentation on the Spotted Owl in the southern California mountains using a SEPM developed specifically for this species (Schumaker 1995). We focus on the spotted owl because: a) the southern California population is small, its habitat is fragmented, and a panel of experts has expressed concern about its long-term viability (Verner at al. 1992); b) it is considered an "umbrella species" for other forest-dependent species and a management indicator species by the Forest Service (Verner et al. 1992); c) an ongoing demographic study in the San Bernardino Mountains provides the habitat-specific information about demography, dispersal behavior, and habitat selection required to parameterize a SEPM (LaHaye et al. 1992); d) prior analyses of spotted owl dynamics in the southern California mountains have not explicitly considered landscape habitat patterns (Noon and McKelvey 1992, LaHaye et al. 1994); and e) the digital vegetation database for the study area, described in the last section, has a set of attributes mapped at a spatial scale (2 ha MMU) that is appropriate for modeling spotted owl habitat suitability.

Detailed studies of spotted owl habitat in the Pacific Northwest (Thomas et al. 1993), the Sierra Nevada and southern California (Verner et al. 1992) have shown vegetation type (forest composition), forest canopy cover, and forest structure (distribution of tree size) to be correlated with habitat suitability. Prior to the mapping effort described above, there was no digital map of southern California forests that estimated each of these attributes. It has been suggested in the literature that the grain of a map of habitat attributes should be no more than 1% of the target species' average home range size (reviewed in Schultz and Joyce 1992). Our vegetation type labels nominally meet this requirement; the minimum polygon size is 2 ha, the median is roughly 4 ha and the mean is approximately 15 ha (Franklin and Woodcock, submitted); the territory size for Spotted Owl in the study area is highly variable, roughly 200-1000 ha depending on elevation (LaHaye et al. 1992). The only other completed vegetation database that covers the study area has a MMU of 100 ha (Davis et al. 1995) and an average polygon size of 600-1000 ha (Franklin and Woodcock, submitted). No other existing regional database has information on conifer and hardwood forest cover for all forest types.

Preliminary Results of Modeling the Distribution of California Spotted Owl Territories in the San Bernardino Mountains

The SEPM we are using partitions the landscape into a uniform array of hexagonal cells, where the hex cell size approximates the average territory size for a pair of owls. The model evaluates the quality of habitat within hex cells based on user-defined scores for each habitat class, and identifies cells which are suitable breeding territories based on a user-defined threshold score. The goal is to select a hex cell size, habitat classes, class scores, and a territory threshold score that will produce a map that closely approximates the true number and spatial configuration of potential owl territories in the landscape. To identify appropriate values for these parameters, we are fortunate to have one southern California mountain range (the San Bernardino Mountains) where the spotted owl population has been intensively studied and a full inventory of owl territories has been completed (LaHaye et al. 1992). What we present here is preliminary work using known owl nest locations from the San Bernardino Mountains to develop reliable habitat suitability parameters for the SEPM.

As previously described, the three mapped vegetation attributes are vegetation type, tree crown size, and percent canopy cover. Going into this analysis, there were several clear trends in owl/habitat relationships that guided our thinking: 1) empirical data shows that spotted owls nest and roost predominately in dense, mature forest stands with large trees (see summary in Verner et al. 1992); and 2) owl territories in low-elevation forest types (bigcone Douglas fir and live oak) are in distinctly smaller patches of forested habitat than territories found in high elevation forest types (mixed conifer and Jeffrey pine). The latter relationship can be seen clearly in Figure 1, where the small low- elevation forest patches in the lower left portion of the map contrast sharply with the continuous forests found at higher elevations. The sizable variability in habitat quantity per territory presents a problem for parameterizing a model that requires a single hex cell size. We addressed this problem by adopting a hex cell size (345 ha) that is skewed towards the lower end of the estimated territory size range (200-900 ha) to increase the relative importance of small forest patches, and assigning higher scores to the low- elevation forest types to compensate for their smaller quantities.



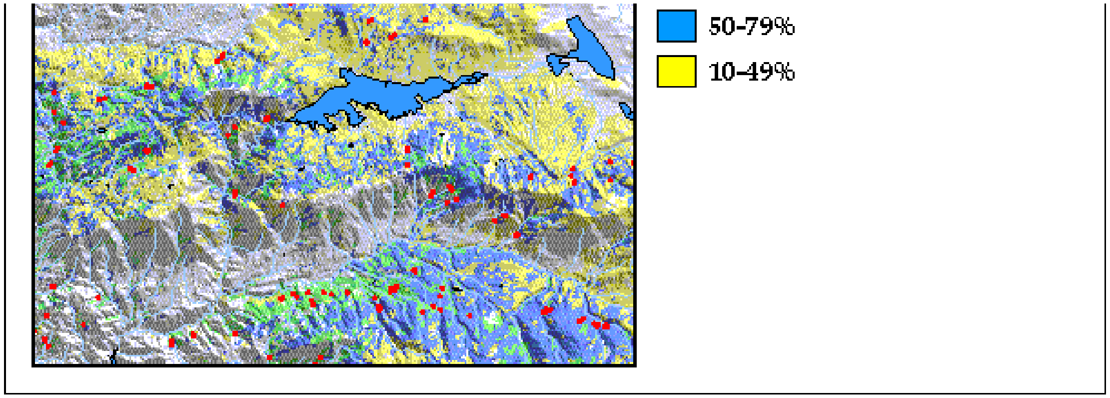


Figure #1. Forest vegetation, cover, and size classes with spotted owl nest sites.

Examinations of the mapped vegetation immediately around owl nest sites indicated that the nest sites were not as highly correlated with high crown size and canopy classes as would be expected from field measurements taken at the nests. Based on the field data, most nests should fall in the 80-100% canopy and 24-40+ crown diameter classes and essentially all of them should be in or above the 50-79% canopy class. What we found for the entire San Bernardino Mountain study area, using the mapped categories, was 48% (147 of 305) of the nest sites in the 80-100%/24-40+ class and 82% (250 of 305) in or above the 50-79% canopy class. Also, 10% (31 of 305) of the nest sites were identified as being within vegetation types that spotted owls do not nest in (chaparral and pinyon pine).

There are several possible explanations for the observed discrepancies between the field data and the map. One is that some stands may simply be mislabeled; there are undoubtedly some errors in the map and a classification accuracy of 82% for nest sites by canopy class is reasonable to expect. Also, preliminary accuracy assessments suggest that the three mapped vegetation attributes (i.e. type, size, and cover) may vary significantly in their accuracy level, with tree size being the least accurate of the three. However, it is also likely that scale differences between the field data and the map is a factor. Field data on forest structure around nest sites was collected using tenth-acre plots, while the map's minimum mapping unit is approximately 5 acres. It is probable that in some areas the forest structure immediately around the nest site is not representative of the larger mapped unit. This is almost certainly a factor in the appearance of nest sites in non-forest types; some low-elevation nests are located in extremely narrow riparian forest stands surrounded by chaparral.

These map scale and accuracy issues were considered when identifying the habitat classes to use in the model. Although we initially started with a good idea of the habitats that are most important to spotted owls at the micro-scale, it was necessary to determine which mapped, stand-scale vegetation classes best correlated with known nest sites. We did this by quantifying the amounts of each vegetation class found within the 345 ha hexagon cells that encompassed known nest sites (Table 1). In this way, we were able to identify which mapped attributes were good predictors of owl nests and which were not. Based on this analysis, we decided to drop tree crown size from the model and focus on vegetation type and canopy cover. Although the highest crown size class correlated well with nests in some areas, it did not correlate well with known nests in oak woodlands. With over 60% of this vegetation type concentrated in a single, mid-range size class, this attribute was not effective at discriminating good habitat in live oak woodland.

Table 1. An example of the differences in habitat quantities around known owl nest sites and how this is reflected in the habitat's assigned "habitat suitability" score. Shown are the dominant vegetation classes by territory type, the range and average percent of that habitat within 345 ha hexagonal "territories" that contain known nest sites, and each class's assigned score for the predictive model. The threshold score for a hex cell to be identifed as an owl territory is 27.

Territory Type Vegetation Classes (by Forest Type and % Canopy Cover)	% Range of this Habitat in 345 ha Cell	Avg. % of this Habitat in 345 ha Cell	Assigned Score
Mixed Conifer Territories			
Mixed Conifer, 80-100%	2 - 51%	24%	55
Mixed Conifer, 50-79%	1 - 88%	37%	25
Mixed Conifer, 10-49%	0 - 70%	15%	10
Mixed Conifer/Jeffrey Pine Territories			
Mixed Conifer within Pine Forest, 80-100%	0 - 24%	5%	50
Mixed Conifer within Pine Forest, 50-79%	5 - 60%	25%	50
Mixed Conifer within Pine Forest, 10-49%	0 - 37%	11%	10
Jeffrey Pine, 50-100%	1 - 52%	19%	25
Jeffrey Pine, 10-49%	6 - 65%	25%	15
Bigcone Douglas Fir/Canyon Live Oak Territories			
Bigcone Douglas Fir, 50-100%	2 - 35%	14%	99
Bigcone Douglas Fir, 10-49%	0 - 8%	2%	45
Live Oak, 80-100%	1 - 12%	6%	80
Chaparral	41 - 97%	73%	8

Possible scores that can be assigned to habitat classes range from 0 to 99. Bigcone Douglas fir with 50-100% canopy cover (the most important low-elevation forest type) received the maximum score of 99. Other vegetation type/canopy cover classes were scored proportionally lower based on their importance as owl habitat, their relative abundance in a territory, and the vegetation types they are associated with. A key factor in scoring habitat classes is the threshold score that a hex cell must exceed in order to be identified as a territory. We selected a threshold score of 27 based on trial and error. This threshold value achieved adequate representation of the low-elevation territories embedded in chaparral. At this threshold and with chaparral habitat scored at 8, a hex cell with 21% bigcone Douglas fir (72.5 ha) surrounded by chaparral would be designated as a territory.

The most important high-elevation forest type, mixed conifer with 80-100% canopy cover, was scored based on its mean quantity (areal extent) in known owl territories relative to mean quantities found in bigcone Douglas-fir territories. On average, it is 56% more abundant than bigcone Douglas-fir, thus its score was calculated to be 56% of 99 (55) (see Table 1). Forest types in the lower canopy cover classes were scaled down based on their considerably lower value as owl habitat. However, an exception to this was in mixed conifer habitats embedded in Jeffrey pine forest. Our analysis found that owl territories in these more arid, rainshadow forests were strongly associated with patches of mixed conifer habitat that have canopy cover of 50-79%. To account for this, we gave higher weighting to this habitat class.

Figure 2 provides a visual representation of how the model-predicted territories correspond with known owl nest sites and Table 2 provides a numeric comparison by both vegetation type and geographic region. The model predicted 131 territories in a mountain range where there are 135 known territories. More importantly, although there is significant error in locational accuracy, the model performed well in estimating total numbers both by vegetation type and by geographic region. Thus, both the size and the spatial distribution of the population are realistically simulated. Clearly, the model isn't perfect. Model-predicted territories are more tightly clustered than actual ones and they tend to overvalue oak woodland habitats in some areas because of difficulties in distinguishing mature stands of trees from dense brushy ones. However, for spatially explicit population modeling, this territory map captures far more spatial and categorical detail about the owl population's distribution than have previous studies (Noon & McKelvey 1992, LaHaye et al. 1994). The real test will be how the model performs when applied to the other southern California mountain ranges.

Predicted Spotted Owl Territories vs Actual Nest Locations

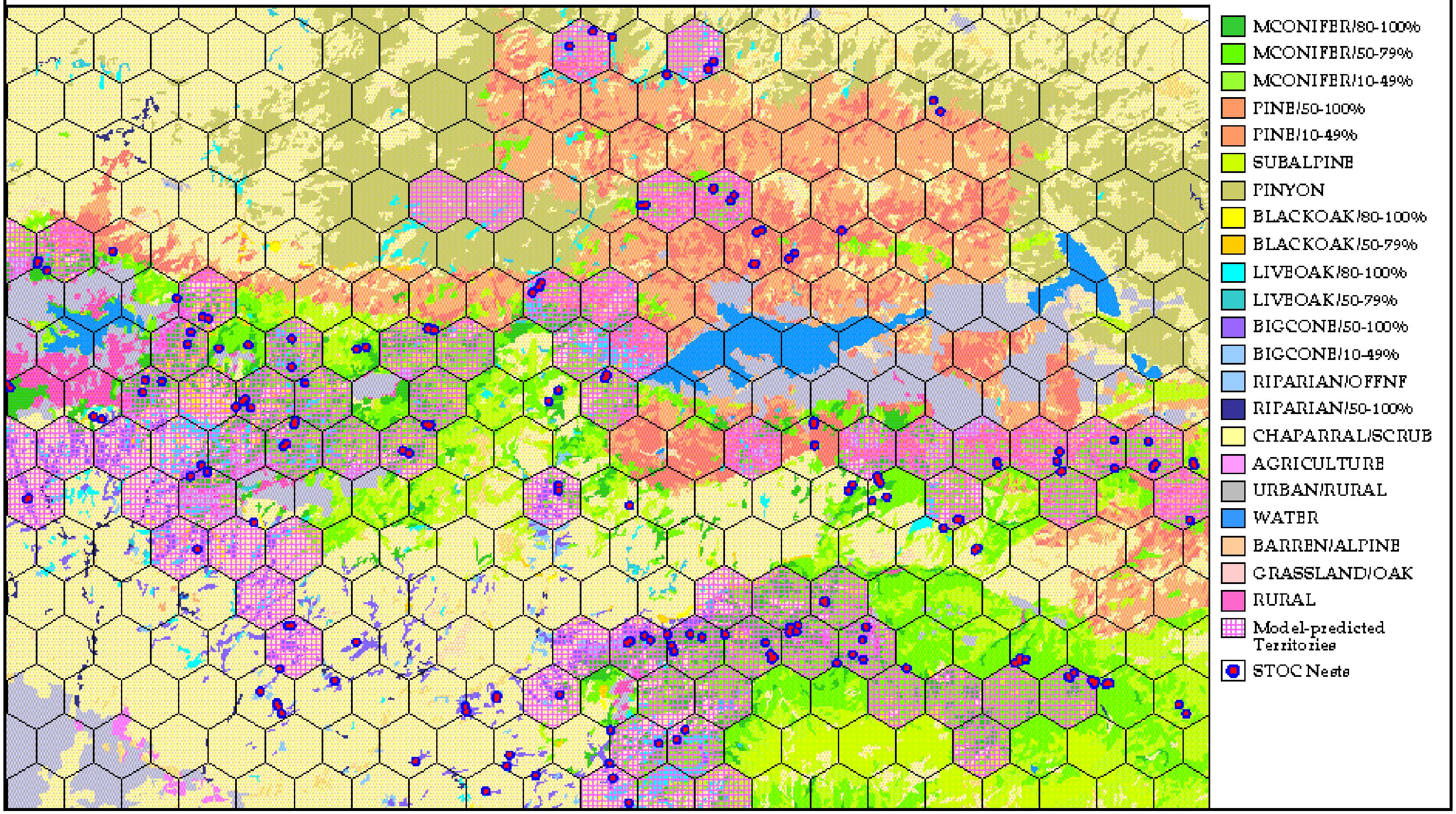


Table 2. Compares the total numbers and locations of model-predicted territories with field data on known spotted owl territories. Comparisions by vegetation type and geographic region are presented. For the model to adequately represent the real owl population it must closely match the true spatial distribution of the population. Although the locational accuracy errors are sizeable, there is generally good correspondence in total numbers, both by vegetation type and geographic region.

	Total Numbers		**Locational Accuracy of Model-Predicted Territories**			
	Model-Predicted Territories	Actual Known Territories	Location Correctly Predicted	Location Not Predicted (Omissions)	Location Predicted, but isn't occupied	%Correctly Predicted
Total Study Area:	**131**	**135**	**88**	**47**	**43**	**65%**
By Vegetation Type:						
Mixed Conifer	43	47	32	15	11	68%
BigconeDF/ Canyon Live Oak	26	29	15	14	11	52%
Mixed Conifer/ BigconeDF/Live Oak	41	30	28	2	13	93%
Mixed Conifer/ Jeffrey Pine	17	20	11	9	6	55%
Live Oak/Pinyon/ Jeffrey Pine	4	6	2	4	2	33%
Jeffrey Pine	0	3	0	3	0	0%
By Geographic Region:						
Northeast	6	9	4	5	2	44%
Big Bear V→Onyx Pk	16	19	11	8	5	58%
AngOaks→HeartBar	21	22	15	7	6	68%
South San Gorgonio	34	24	19	5	15	79%
Lower Front Country	19	23	10	13	9	43%
Arrowhead L. Region	19	23	16	7	3	70%
Silverwood L. Region	16	15	13	2	3	87%

Discussion and Conclusions

We have implemented a suit of mapping methods, relying on digital image processing and the integration of remote sensing with GIS, to estimate vegetation attributes (composition and structure) at a fine grain (2-20 ha stands) for a large region (2 m ha). This paper demonstrates that the resulting database contains a necessary set of attributes at a resolution sufficient for modeling the distribution of suitable habitat for the California Spotted Owl, a forest-dependent umbrella species whose numbers are declining. The habitat-specific, spatially explicit population modeling that is currently in progress will hopefully provide results that indicate the effect of habitat fragmentation on population viability, and will identify habitat that is critical, on the basis of its location, for the viability of the metapopulation.

The lack of perfect correspondence between modeled and actual territory locations, based on the habitat suitability model, could reflect inaccuracy or imprecision in the mapped attributes as discussed above. We are currently conducting an analysis of the sensitivity of the habitat suitability model to error in the attribute labels (Ellen Hines, Masters thesis in progress). It might be possible, using photointerpretation- or field-based editing, to improve the accuracy of the variables that are key to a particular application; our mapping methods lend themselves to this kind of updating. Another possibility is that habitat variables important to spotted owls are not mapped as attributes, and may be difficult to capture in a large-area map. Examples would be local and subtle variations in canopy structure or microclimate, affecting prey populations. This has been discussed in the literature on GIS-based habitat suitability modeling. The forest structure attribute that is most difficult to estimate from remote sensing, or from field data for that matter, is tree crown size. We are exploring the use of spatial variance in high-resolution SPOT panchromatic imagery (10 m pixels) to better estimate forest canopy structure (David Shaari, Masters thesis in progress).

This southern California database has also been used for other environmental modeling applications. Jennifer Swenson (1995) simulated future urban development in the region surrounding the Santa Monica Mountains National Recreation Area (administered by the National Parks Service), and examined its potential effect on the connectivity of the remnant natural vegetation. The results predicted a disproportionate impact of potential future urbanization on certain sensitive vegetation types (oak woodland and coastal sage scrub) primarily due to their proximity to current urban land use. Modeling also indicated that patterns of future urban development are likely to increase the fragmentation of the remaining wildlands, due to the patterns of land ownership in the Santa Monica Mountains.

We hope, in the future, to use predictive vegetation modeling (Franklin 1995) to add detail on vegetation composition to the database. We have tested this approach, with some limited success, in the Cleveland National Forest in San Diego County. Chaparral series were predicted from satellite and terrain data using classification tree modeling (McCullough 1994), and riparian vegetation types were also predictively mapped from terrain variables using a classification tree (Gray 1994). While the accuracy of predictive mapping was low relative to thematic map accuracy standards, the predicted patterns made sense. We feel there may have been problems with the field data used for calibration and validation but collected for other purposes, including ambiguous labeling of vegetation type and poor registration with the digital database. However, being able to predictively map the distribution of dominant plant species at this scale with reasonable confidence would allow us to model the spatial dynamics of the vegetated landscape based on characteristics of the disturbance regime ,and the life history traits of those plant species (see, for example, Mladenoff in press). This, in turn, could be tied back in to a SEPM to examine in a realistic way how a landscape altered by environmental change or land management practices might affect a management-indicator species.

Acknowledgments

We would like to thank the Greg Nichols, Nathan Schumaker, and Joseph Shandley for their contributions to this research.

References

Davis, F.W., Stine, P.A., Stoms, D.M., Borchert, M.I. and Hollander, A.D. (1995) Gap analysis of the actual vegetation of California 1. The southwestern region. *Madrono* 42: 40-78.

Doak, D.F. and Mills, L.S. (1994) A useful role for theory in conservation. *Ecology* 75: 615-626.

Dunning, J.B., Jr., Stewart, D.J., Danielson, B.J., Noon, B.R., Root, T.L., Lamberson, R.H., and Stevens, E.E. (1995) Spatially explicit population models: current forms and future uses. *Ecological Applications* 5: 3-11.

Franklin, J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19:494-519.

Franklin, J., Logan, T., Woodcock, C. E., and Strahler, A. H. (1986) Coniferous forest classification and inventory using Landsat and digital terrain data, *IEEE Transactions on Geoscience and Remote Sensing* GE-24, 139.

Franklin, J. and Woodcock, C.E., (submitted) Multiscale vegetation data for the mountains of Southern California: spatial and categorical resolution, in, *Scaling of Remotely Sensed Data for GIS,* D. A. Quattrochi and M. F. Goodchild, eds., Lewis Publishers.

Franklin, J. F. (1993) Preserving biodiversity: species, ecosystems, or landscapes?, *Ecological Applications* 3: 202-205.

Franklin, J. F. (1995) Why Link Species Conservation, Environmental Protection, and Resource Management?, in *Linking Species and Ecosystems*, Jones, C. G. and Lawton, J. H., Eds., New York: Chapman & Hall. pp. 326-335.

Goodchild, M. F. (1992) Geographical data modeling, *Computers & Geosciences* 18: 401-408.

Goodchild, M. F. (1994) Integrating GIS and remote sensing for vegetation analysis and modeling: methodological issues. *Journal of Vegetation Science* 5: 615-626.

Gopal, S. and Woodcock, C. E. (1994) Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing* 60: 181-188.

Gray, C.A. (1995) Predicting the location of riparian vegetation using Landsat TM and digital terrain data in the Cleveland National Forest. Masters Thesis, Department of Geography, San Diego State University, San Diego CA.

Henderson-Sellers, A. (1994) Global terrestrial vegetation 'prediction': the use and abuse of climate and application models. *Progress in Physical Geography* 18: 209-246.

James, M., and Kalluri, S.N.V. (1994) The Pathfinder AVHRR data set: an improved coarse resolution data set for terrestrial monitoring. *International Journal of Remote Sensing* 15: 3347-3363.

LaHaye, W.S., Gutierrez, R.J., and Call, D.R. (1992) Demography of an insular population of spotted owls (*Strix occidentalis occidentalis*). In: McCullough, D., Barrett, R.H., eds. *Wildlife 2001: Populations.* New York: Elsevier Press; pp. 803-814.

LaHaye, W.S., Gutierrez, R.J., and Akcakaya, H.R. (1994) Spotted owl metapopulation dynamics in southern California. *Journal of Animal Ecology* 63: 775-785.

Loveland, T., Merchant, J., Ohlen, D., and Brown, J. (1991) Development of a land-cover characteristics database for the coterminous United States. *Photogrammetric Engineering and Remote Sensing* 57: 1453-1463.

Maiden, M. (1994) The Pathfinder project. *International Journal of Remote Sensing* 15: 3333-3345.

McCullough, P.E. (1995) Predictive vegetation mapping using classification tree analysis. Masters Thesis, Department of Geography, San Diego State University, San Diego CA.

McKelvey, K., Noon, B.R., and Lamberson, R.H. (1993) Conservation planning for species occupying fragmented landscapes: the case of the northern spotted owl. In: Kareiva, P.M., Kingsolver, J.G., and Huey, R.B., eds. *Biotic Interactions and Global Change.* Sunderland, MA: Sinauer Press. pp. 424- 450.

Mladenoff, D.J., Host, G.E., Boeder, J., and Crow, T.R. (in press) LANDIS: A spatial model of forest landscape disturbance, succession, and management. In, Goodchild, M., Steyaert, L.L., Parks, B.O., Crane, M.P., Johnston, C.A., Maidment, D.R. and Glendenning, S., editors, *GIS and environmental modeling: Progress and research issues,* Ft. Collins, CO: GIS World, Inc.

Noon, B.R. and McKelvey, K.S. (1992) Stability properties of the spotted owl metapopulation in southern California. In: Verner, J., McKelvey, K.S., Noon, B.R., Gutierrez, R.J., Gould, G.I., Beck, T.W., tech. coordinators. *The California Spotted Owl: a technical assessment of its current status.* General Technical Report PSW-GTR-133. Pacific Southwest Research Station, Forest Service. U.S. Department of Agriculture. pp. 187-206.

Pulliam, H.R. (1988) Sources, sinks, and population regulation. *American Naturalist* 132: 652-661.

Pulliam, H.R., Dunning, J.B. Jr., and Liu, J. (1992) Population dynamics in complex landscapes: a case study. *Ecological Applications* 2: 165-177.

Running. S.W., and 13 other authors (1994) Terrestrial remote sensing science and algorithms planned for EOS/MODIS. *International Journal of Remote Sensing* 15:3587-3620.

Sample, V. A., Ed., (1994) *Remote Sensing and GIS in Ecosystem Management.* Washington, DC: Island Press.

Schultz, T.T. and Joyce, L.A. (1992) A spatial application of a marten habitat model. *Wildlife Society Bulletin* 20:74-83.

Schumaker, N.H. (1995) Habitat connectivity and spotted owl population dynamics. Ph.D. Dissertation. University of Washington: College of Forest Science.

Schumaker, N.H. (Unpubl. manuscript) A simulation approach to quantifying landscape quality.

Scott, J.M. and 11 other authors (1993) Gap analysis: a geographical approach to protection of biological diversity. *Wildlife Monographs* 123:1-41.

Soule, M.E., Bolger, D.T., Alberts, A.C., Sauvajot, R.S., Wright, J., Sorice, M., and Hill, S. (1988) Reconstructed dynamics of rapid extinctions of chaparral-requiring birds in urban habitat islands. *Conservation Biology* 2: 75-92.

Strahler, A. H., (1981) Stratification of natural vegetation for forest and rangeland inventory using Landsat digital imagery and collateral data. *International Journal of Remote Sensing* 2: 15-41.

Strahler, A. H., Woodcock, C. E., and Smith, J. A. (1986) On the nature of models in remote sensing. *Remote Sensing of Environment* 20: 121-139.

Swenson, J. (1995) Habitat fragmentation in the Santa Monica Mountains: current status and future predictions. Masters Thesis, Department of Geography, San Diego State University, San Diego CA.

Thomas, J.W., Forsman, E.D., Lint, J.B., Meslow, E.C., Noon, B.R., and Verner, J. (1990) *A conservation strategy for the Northern Spotted Owl.* Interagency Scientific Committee to Address the Conservation of the Northern Spotted Owl (USDA: Forest Service, USDI: Bureau of Land Management, Fish and Wildlife Service, and National Park Service). 1990-791-171/20026. United States Government Printing Office, Washington, D.C. USA.

Townshend, J.R.G., Justice, C., Li, W., Guerney, C., and McManus, J. (1991) Global land cover classification by remote sensing: present capabilities and future possibilities. *Remote Sensing of Environment* 35: 243-255.

Verner, J., McKelvey, K.S., Noon, B.R., Gutierrez, R.J., Gould, G.I., Beck, T.W., Technical Coordinators. (1992) *The California spotted owl: a technical assessment of its current status.* General Technical Report PSW-GTR-133. Pacific Southwest Research Station, Forest Service. U.S. Department of Agriculture.

White, J. D. and Running, S. W. (1994) Testing scale dependent assumptions in regional ecosystem simulations. *Journal of Vegetation Science* 5: 687- 702.

Woodcock, C. D., Collins, J., Gopal, S., Jakabhazy, V. D., Li, X., Macomber, S., Ryherd, S., Harward, V. J., Levitan, J., Wu, Y., and Warbington, R. (1994a) Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sensing of Environment* 50: 240-254.

Woodcock, C. W., Gopal, S., Macomber, S., and Jakabhazy, V. (1994b) Accuracy Assessment of the Vegetation Map of the Plumas National Forest, Boston: Boston University Center for Remote Sensing, 18 pp.

Woodcock, C. E. and Harward, V. J. (1992) Nested-hierarchical scene models an image segmentation. *International Journal of Remote Sensing* 13: 3167-3187.

Janet Franklin
Department of Geography
San Diego State University
San Diego CA 98182-4493
Telephone: (619) 594-5491
FAX: (619) 594 4938
Email: janet.franklin@sdsu.edu

John Stephenson
Cleveland National Forest
USDA Forest Service
10845 Rancho Bernardo Rd., Suite 200
Rancho Bernardo, CA 92127-2197
Telephone: (619) 674-2951
Email: jstephen@typhoon.sdsu.edu

Integrating Stratified Sampling, Canonical Correspondence Analysis, and GIS for Predictive Vegetation Modeling in the Spring Mts. of Southern Nevada.

Andrew D. Weiss (1)
Stuart B. Weiss (1)
Alisya T. Galo (1)
Jan Nachlinger (2)
Daniel Pritchett (2)

1. Center for Conservation biology
   Dept. of Biological Sciences
   Stanford University
   Stanford CA 94305

2. The Nature Conservancy
   Northern Nevada Project Office
   1885 S. Arlington Ave. Suite 1
   Reno NV 89509

Canonical Correspondence Analysis (CCA) and workstation Arc/Info GIS are used to analyze vegetation distributions in the Spring Mountains of Southern Nevada. These mountains range in elevation from 700m - 3600m in an area of 1286 square kms, and exhibit diverse plant communities including Mojave desert scrub at the base, several shrub and forest communities at intermediate elevations, culminating with high elevation Bristlecone pines and alpine meadows. The vegetation dataset consists of 230 plots selected through a GIS-based stratified sampling design incorporating physiographic and geological variables.

CCA generates ordination axes that are linear combinations of environmental variables, and calculates the centroids and tolerances of the species or communities within ordination space. GIS is used to project the values of the ordination axes across geographic space, and to classify the landscape into probability or abundance surfaces for each species or community.

Based on preliminary accuracy assessment, the predictive maps accurately show both coarse scale zonation by elevation, and finer scale variation based on topographic position and insolation gradients. Accuracy assessment is continuing with additional fieldwork, satellite imagery, and comparison with the Nevada GAP analysis vegetation maps.

Estimating Spatial Uncertainty as a Function of Scale: Implications for Landscape Ecology

Carolyn Hunsaker, C.R. Ehlschlaeger, F.W. Davis, M.F. Goodchild

ABSTRACT:

Maps are representations of reality, and they contain errors that are spatially autocorrelated. Land cover is visualized using categorical coverage maps where data quality is determined by whether the category labeled at each geographic location in the database matches what really exists on the ground. This research begins to address the implications of spatial uncertainty that are a function of map scale for common landscape ecology applications.

From a truth map, autocorrelation properties--the maximum distance of spatial dependence and the distance decay exponent--are developed for each land cover class. Confusion matrices (error matrices or contingency tables) for proportions of cover classes are used to better understand categorical data for the maps being compared. Information from the confusion matrix and the autocorrelation properties of the truth map are then used to generate potential realizations of a coarser resolution map. The study was performed on a 1000 by 1000-cell sample of the truth map using three land cover classes: California coastal sage scrub, urban, and other. To estimate the influence of spatial error on ecological applications we apply a habitat suitability model for an endangered bird that is dependent on the coastal sage scrub class, the California gnatcatcher, and some commonly used landscape pattern metrics. The habitat model is based on patch size, patch density, and distance between patches. The metrics include patch shape complexity, dominance, contagion, and amount of edge between cover classes.

Carolyn T. Hunsaker
Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6038

C.R. Ehlschlaeger, F.W. Davis, M.F. Goodchild
University of California
Department of Geography
Santa Barbara, CA 93106-4060

EOS Potential User Model Development

John E. Estes

This paper documents the conclusions of the Earth Observing System Data and Information System (EOSDIS) Potential Users Conference.

EOSDIS provides and will continue to provide the basic means of accessing the data and informati on products of NASA's Mission to Planet Earth (MTPE), including the Earth Observing System (EOS). EOS is being implemented to acquire global data by sensor systems operating at wavelengths across the electromagnetic spectrum. EOSDIS is a distributed data and information system that is being implemented to help the user community acquire the data and information they require. NASA's MTPE, is a component of the U.S. Global Change Research Program, which was initiated under President Bush to understand th e interactive physical, biological and socioeconomic processes that regulate the total Earth system and to predict how this system will change in response to natural forces and human actions. Mission to Planet Earth includes EOS, a set of EOS precursor and Earth Probe satellite missions, NASA's Earth Science Research and Analysis Program, and EOSDIS. Thus, the products of MTPE will consist primarily of sets of satellite remote sensing data and the diverse information. Within these efforts, EOSDIS is NASA s system that is essential to the U.S. Global Change Research Program (USGCRP) in serving both the science and the larger needs of society for global change information.

EOSDIS is being designed and implemented as a mechanism for generating, archiving and distributing the products of Mission to Planet Earth to the full range of the user community (i.e., from intermediate to end users). Until recently most users of EOSDIS had been from the natural science portions of the global change research community. Th is has been the community that has by-in-large provided those users who: have served on various committees advising system developers and operators; and, who have tested prototypes of system capabilities prior to their general release for use. There are no intrinsic limits to the interests of users that can be served by EOSDIS; but, there will inevitably be limits to the capacity of EOSDIS to provide a user with choices relative to the style and type of services he or she may choose. Given this recognit io n, what is the nature of the potential set of EOSDIS users, what characteristics should the system have if it is to serve them, and how many such users could there be? Providing authoritative answers to these questions is the objective of this report. O nce answers are in hand, one can reasonably ask what portion of the set of services potential users require can be served within the limits of government's financial investment? A complimentary issue which must be addressed and which will evolve over tim e is the essential set of information resident within EOSDIS that is needed to serve science and the far broader needs of society as well.

NASA is implementing EOSDIS using an easily-extensible, evolvable, logically distributed, open architecture. This architecture allows distribution of EOSDIS functional elements to various physical locations to take advantage of institutional capabilities and science expertise. These same attributes allow EOSDIS and other information systems to be made to appear to users as one virtual or logical system wherein data can be searched, browsed, ordered and delivered. Regardless of the physical distribution of the elements of EOSDIS, the system will appear to be a single logical entity to a given user; although performance w ill be subject to the vagaries of network performance over which EOSDIS will have no control. The key aspect of EOSDIS architecture is its middle ware or interfaces which will allow a variety of diverse logical or physical service providers to respond to a range of different clients.

Within this architecture or framework of interactions, EOSDIS will provide a set of services covering the ingest, processing, archiving, information management, and distribution of its data and information holdings. In addition, EOSDIS will provide an ad vertising service allowing others to make known the services they offer to any user of EOSDIS and at least one version of a client which uses the standards and approaches of EOSDIS' architecture to obtain services from EOSDIS including searching, ordering, and browsing. Accounting and security services are imbedded in the lower level protocols which form the basis for the middle ware. Figure 1 shows one view of this architecture.

The overall flow of data from satellites through various networks to archives and subsequently to users is illustrated in Figure 1. For purposes of understanding and describing user requirements, there are several key aspects of this flow. First, some of the data from EOS satellites and Landsat-7 will be directly downlinked or broadcast to users. EOSDIS will make available the software that is needed to process these data so that users with requirements for extremely timely access to data can use them in the implementation of their own data processing and delivery systems. Second, EOSDIS is providing to NOAA an ability to tap the EOS data stream where it reaches the ground. Operational users of EOS data can work with NOAA to obtain data through this means within three hours of data acquisition. Third, users may obtain data from EOSDIS archive and distribution elements with timeliness being a function of the level of processing of the product and the need for quality control of products, as follows.

- Level O-Reconstructed unprocessed instrument/payload data at full resolution; any and all communications artifacts (e.g., synchronizations frames, communications headers removed.
- Level 1A - Reconstructed unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary i nformation, including radiometric and geometric calibration coefficients and georeferencing parameters (i.e., platform ephemeris) computed and appended but not applied, to the level 0 data.
- Level 1B - Level 1A data that have been processed to sensor units (not all instruments will have a level 1B equivalent).
- Level 2 - Derived geophysical variables at the same resolution and location as the level 1 source data.
- Level 3 - Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.
- Level 4 - Model output or results from analyses of lower level data (i.e., variables derived from multiple measurements).

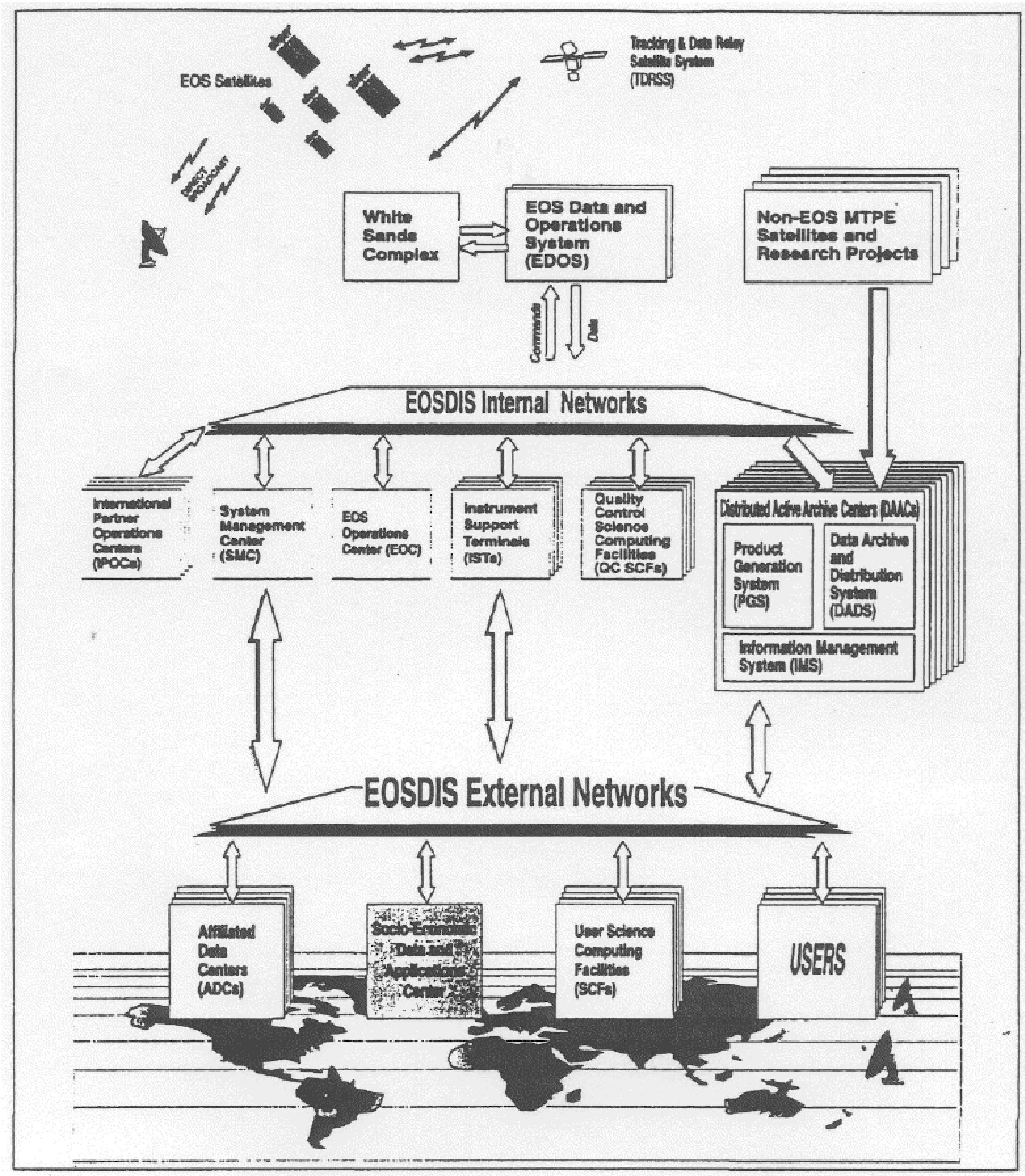
Nothing in the system architecture precludes the addition of special taps on the data and information flow, but the costs associated with such extra capabilities are not included in the current EOSDIS budget or plans.

From the user standpoint, the key elements of the present EOSDIS implementation are the set of nine Distributed Active Archive Centers (DAACs). At mo st of these centers, the user services and support functions of EOSDIS are collocated with the information technology functions of data ingest, processing, archiving, distribution and information management. These DAACs are the primary entry points for t he user community. Each is located at an institution with a commitment to conduct research which includes use of data from the local DAAC.

Four DAACs are located at NASA Centers. These DAACs and the areas of special emphasis of their personnel in Earth sc ience are: Goddard Space Flight Center (GSFC), upper atmosphere, atmospheric dynamics, global biosphere and geophysics; Jet Propulsion Laboratory (JPL), ocean circulation and air sea interaction; Langley Research Center (LaRC), radiation budget, aerosols, and tropospheric chemistry; and, Marshall Space Flight Center (MSFC), hydrology.

The remaining five DAACs are located at universities or non-NASA U.S. Government agency facilities. These data centers and the areas of special emphasis in Earth science of t heir personnel areas follows. The Alaska SAR (Synthetic Aperture Radar) Facility (ASF), at the University of Alaska, Fairbanks, Alaska. The areas of emphasis of the ASF, DAAC are: sea ice, and polar process imagery. The U.S.Geological Survey (USGS), E arth Resource Observation Satellite (EROS) Data Center (EDC), at Sioux Falls, South Dakota. The area of emphasis of the EDC, DAAC is: land processes imagery. The National Snow and Ice Data Center (NSIDC) at the University of Colorado, Boulder, Colorado. The area of emphasis of the NSIDC is: non-SAR cryosphere data. The Department of Energy (DOE), Oak Ridge National Laboratory (ORNL), Oak Ridge, Tennessee. The area of emphasis of the ORNL, DAAC is: ground-based data relating to biogeochemical dynamics. The fifth non-NASA data center is the Socioeconomic Data and Applications Center (SEDAC) at the Consortium for International Earth Science Information Networks (CIESIN) at Saginaw, Michigan. The area of emphasis of the SEDAC is: social science and econo mics applications. An extended discussion of these EOSDIS elementscan be found in the 1993 EOS Reference Handbook (NASA 1993).

Figure 1-1. EOSDIS Networks



A great deal of planning and development activity has occurred in the EOS program since it was given it's Presidential Initiative "new start" in October 1990. Initial EOS investigations and instruments were selected in 1989. The EOSDIS Core System con tractor was selected in 1992. It is understood that user needs for EOSDIS will become better under stood as users work with and respond to current versions of the system. EOSDIS is being designed to maximize openness to change through the incorporation of layering and standards and vendor independence in the development and prototyping of the system. EOSDIS evolution has already begun and will proceed through a series of major versions that are described below.

EOSDIS Version 0. Starting with existing heterogeneous, distributed Earth science data systems, Version 0 is in the process of evolving toward s EOSDIS by taking maximum advantage of existing experience and by ensuring that no disruption in current user services occurs. This system has become operational within the last year and supports the scientific community in day-to-day research activitie s. Through the interconnection of existing systems, Version 0 currently serves as a functional prototype of selected key EOSDIS services. As a prototype Version 0 does not posses all the functional capabilities, fault tolerance, or the planned reliabili ty of later versions. These uses test existing services to determine what additional or alternative capabilities are required of the fully operational EOSDIS. Version 0 interconnects existing Earth science data systems via electronic networks, interoperab le catalogs, and common data distribution procedures to provide better access to existing and pre-EOS data.

EOSDIS Version 1. Version 1 of EOSDIS is scheduled to be implemented in steps over the period 1995-1997. Version 1 will be designed and developed while Version 0 is operating. As such the developers of this version of EOSDIS will benefit from Version 0 user feedback. As currently envisioned, Version 1 will be physically distributed, but the system will appear completely integrated to users. It wi ll provide researchers, in particular, with the complete set of capabilities needed for EOS science and mission operations. The new capabilities of the Version 1 data and information system, after a period of parallel operations, will replace the appropr iate elements of the Version 0 system. This transition to enhanced user services between systems will occur in as smooth and as transparent as practical.

EOSDIS Version 2. EOSDIS Version 2 will be implemented prior to the launch of the first EOS satellite in 1998. As currently planned, the mission and instrument command and control and product generation functions, discussed in the earlier versions above, will be greatly enhanced. The full, planned capacity of EOSDIS will be available to support all of the functions required of the first EOS platform.

Subsequent versions of EOSDIS will supplement capacity and services as required by EOS spacecraft launches. EOSDIS capabilities will continue to be evaluated and upgraded to meet the needs of its user community, and technology will be upgraded and/or augmented as needs arise and are judged appropriate. Clearly, the needs of potential EOSDIS users identified in this report would have to be phased into EOSDIS as it evolves through this series of versions. Th e present report is intended to initiate and provide a foundation for addressing the full range of potential uses for EOSDIS and its information holdings.

In serving users, the major role for EOSDIS is to provide archival and distribution services for NASA's Earth science data, related data from other sources and information products derived from these data. As such, EOSDIS data sources include: digital data from past missions, new products generated from existing datasets, data from EOS precursor and Eart h Probe satellite missions, standard data products of EOS instruments, selected data generated by Earth remote sensing missions of NASA's International Partners and by ground-based studies conducted through the research and analysis activities of MTPE. At the highest level of abstraction, EOSDIS provides access today to over 400 datasets. This total is expected to grow to over 800 datasets over the next decade. Many of these datasets will be quite large with individual users only desiring access to small subsets. In some cases, a few users may access some lower level, less easily interpreted datasets. Use of the products derived from these datasets will, however, be predicated upon the ultimate availability of these data from an accessible, authoritativ e, well maintained archive. For the purposes of this report the content of EOSDIS is deemed fixed, determined by the research priorities of MTPE and inherent in the choices of instruments flown in orbit.

EOSDIS can provide a variety of other services in addition to archiving and distributing Earth science data. EOSDIS could also if desired generate user-defined products involving the combination and manipulation of data from all sources as well as their incorporation into models of the environment. Today, EOSDIS provides the ability to perform system wide searches by data type, location, and time. Metadata (data about data) are associated with all EOSDIS holdings, and these metadata are being upgraded to comply with the spatial (georeferenced) metadata sta ndards as part of the development of the National Spatial Data Infrastructure mandated by Executive Order 10926.

Through EOSDIS, the user community can also have access to data held by both other United States Government and International Partner agencies. As discussed in Chapter 2, the interconnection of the data and information systems of the government agencies associated with the U.S. Global Change Research Program (USGCRP) is currently known as the Global Change Data and Information System (GCDIS) and consists of gopher and Mosaic servers available on the Internet. A similar concept for establishing an international environmental information infrastructure or digital library is being pursued through various cooperative means, especially through ini tiatives under the G-7 or Economic Summit. Both concepts, GCDIS and the international environmental information infrastructure, are based on establishing virtual information systems through high-level interconnections of independently implemented systems. There are many other potential services which could be provided in association with EOSDIS' information holdings. Which services do potential users feel they will need and use? Again, providing a good current answer to this question is also part of this report. With this answer as a basis, the program can begin to determine which of these services should be provided by EOSDIS directly to users and which should be provided by others who extend EOSDIS and may provide value-added or derived products as well.

The community of current and potential users of environmental remote sensing data and research results is extensive and varied. In this report, to adequately represent this diversity, the potential user community was divided into broad somewhat overlapping segments. These groupings were not intended to be exclusive as it was realized that some users could function in more than one way. A chair and chapter writing committee was selected for each group. Participants were selected by the chapter chairs in concert with the overall Potential User Model Development co-chairs and are intended to constitute a representative set of authorities in the various areas covered. Accommodation of personal schedules, willingness to devote time to this activity and the usual exigencies of reaching individuals have all played a role in this as well. As an authoritative statement, this report is no better than its authors and we offer it as a reasonable statement of the current view of the potential user community of EOSDIS.

Chapter chairs together with the Potential User Model Development effort co-chairs constitute the overall editorial committee for this document. Work on this report began in December, 1994 and the conference was convened during the week of June 19, 1995.

The universe of potential EOSDIS users was divided up into twelve (12) segments. The first four groups cover activities of the global change or environmental research community which NASA has been working with in the development of EOS. This community has been subdivided into groups which include three aspects of the conduct of research, and the production of scientific assessments of the state of our knowledge in this policy relevant research arena. Chapter 3 focuses on those who could extend EOSDIS and add to the suite of services and information products provided by the research community. Chapter 4 expresses the needs of those who could use EOSDIS' data and information to support government operations including the provision of operational information products. Chapters 9 through 13 cover the aspects of dealing with the environment, e.g., resource planning and management, the formulation of policy including treaties, laws and regulations, and decision making; and the use of EOSDIS data products by the legal community in the enforcement and adjudication of established policies. The education community has been divided into two groups -- education for non-specialists at the primary, secondary and introductory college levels (K-14) and professional education including upper level college courses, graduate level and continuing professional training. Chapter 14 covers the daunting task of dealing with the needs of the general public and those who directly provide/serve them information, namely the library community and the press. General descriptions of the individual groups are as follows:

1. Retrospective Research
   The scope of this chapter includes environment and natural resources research and social science issues that primarily use existing data products including new data and information being provided by satellite and non-satellite sources.

2. Field Campaigns and Individual Data Providers
   The focus of this chapter is environment and natural resources research activities which acquire data in an intermittent mode such as through major or minor field campaigns such as the Tropical Ocean Global Atmosphere (TOGA) experiment, the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE), as-well-as the work of individuals or groups for the duration of a one-time contract or grant.

3. Persistent Information Production for Research
   This chapter focuses on those individuals, agencies and organizations who develop information on a routine basis from continuously collected raw data or from derived information based on such data (i.e., the development and maintenance of research and operational algorithms, and model assimilation of data) for primarily research purposes.

4. Scientific Environmental Assessment
   The needs for EOSDIS support to those involved in the process of producing major environmental assessment reports (e.g., the generation of the Intergovernmental Panel on Climate Control (IPCC), and ozone reports) is the subject of this chapter.

5. Commercial Value Added Service Providers
   This chapter targets commercial activities which provide value added services. This includes individuals and private entities, both for-profit and not-for-profit, that could access or extend EOSDIS and subsequently derive data and information products and/or enhanced data access and services for their clients.

6. Operational Users
   Those individuals or organizations that routinely conduct remote sensing based environmental analyses or generate environmental products (e.g., maps, charts, forecasts, digital datasets, etc.) for use in nearly every sector of society, from policy makers in the U.S. Congress to the general public.

7. Resource Planners and Managers
   The scope of this chapter includes all those who are involved in the planning, use, consumption, or management of any natural or environmental resource (e.g. farming, fisheries management, rangeland management, running a national park). This includes those who do such work as part of their business, as a fiduciary trust, and/or as part of a governmental function.

8. Policy Formulation and Decision Making
   This chapter is devoted to assessing the characteristics of the diverse policy and decision-making communities at local, national, and international levels, who are potential end users of EOSDIS and of knowledge derived from EOSDIS.

9. Use of EOSDIS Data and Information Products by the Legal Community
   This chapter includes the needs of the user community that is involved in the enforcement of environmental regulations, and the use of environmental data in the legal system.

10. K-14 Educators and Students
    This chapter is intended to cover users in the traditional K-12 community and the introductory and survey courses that are given at colleges and universities which are typically take n by a broad range of students who are pursuing degrees in a wide range of majors.

11. Collegiate and Professional Education
    Included in this chapter are those who are involved in teaching upper level undergraduate courses taken as part of specific training in an individual major or closely related group of majors, graduate education. Continuing education for professionals making significant use of environmental data is also included here.

12. Libraries, the Press, and the Public
    This chapter address es uses of EOSDIS in the most general sense by those who typically interact directly with the public at large. An attempt is also made here to address the issues involved in providing EOSDIS data and information directly to the public for non-specific purposes.

In order to address the questions described earlier, each of the twelve groups were asked to discuss five key characteristics of their segment of the potential user community; where possible responses were to be given as a function of time for the period 1995 to 2001. These five key characteristics and their definition, for the purposes of this document, are:

1. User population. Depending on the user group, this may appear as the total number of individuals, organizations, or both; and may include numbers for significant subgroups within a particular segment of the user community. This characteristic is deemed important in order to better understand the total number of probable and possible users.

2. Intensity of use. This characteristic is intende d to cover the frequency and duration of sessions for individual users. Also covered here are the projected percentage of the use which will involve standing orders or machine to machine automated interactions and the anticipated response time of the use rs during their interactions. For example, some users may wish to have an E-mail style response while others envision use through interactive sessions with a human user actually sitting at a terminal or work station.

3. User resources. This covers the resources that typical users within a segment of the community will bring to interactions with EOSDIS including experience, sophistication in information system use, intellectual ability and education, financial resources, software, hardware, and connectivit y. If the group has a broad range of available resources, this characteristic should include an indication of the distribution of the user population across this range.

4. User activities. This characteristic covers the actual tasks or work that the users would like to accomplish using EOSDIS. The focus should be limited to those things which users do today or which the new data sources to be held in EOSDIS will enable. Speculative new uses are to be avoided as they are likely to be difficult to characterize reliably.

5. Desired EOSDIS functionality and capabilities. This characteristic covers the projected products and services that EOSDIS or its extensions must have in order to serve the specific community, enhancements to currently planned contents and performance to enable, enhance or facilitate the user activities described above. Enhancements to currently planned content is intended to include additional data products which can be readily derived from the

contents of EOSDIS. Although recommendations for achieving interoperability/compatibility with specific other data and information systems and services is appropriate, recommendations that NASA should collect different data are not appropriate to the purposes of this document.

It should be noted here that not all potential user groups responded explicit to the five questions posed. There are a number of reasons for this including the various levels of interaction and pre-existing knowledge of the EOS program of the participants. A number of these potential user groups represent sources of latent rather than already expressed demand for the types of data that will be found in EOSDIS. It was felt the detailed characteristics of these user groups (e.g., those users in the Policy Formulation and Decision Making Area) cannot be anticipated fully in advance, and this can only begin to provide a basis for developing detailed information system requirements. This inevitably poses the challenge of how can the designers of EOSDIS maintain the ability to respond in a flexible manner to diverse user communities that could grow quite rapidly as applications are demonstrated.

CONCLUSIONS AND RECOMMENDATIONS

Based upon the work at the conference, a number of conclusions and recommendations have been formulated. These conclusions and recommendations range from the general to the specific. Included here are those general recommendations considered significant by all participating groups. Those more particular conclusions and recommendations repeated by more than one potential user group and chapter specific conclusions and recommendations are not included by the sake of brevity. Those individuals wishing more information can contact the author for a copy of the report which we expect out in the May 1996 time frame.

GENERAL CONCLUSIONS

The general conclusions are:

- The EOSDIS potential user community is large and diverse, and has many shared values and needs.
- Although EOSDIS was designed and is being implemented primarily in support of the Global Change Research community, EOSDIS can potentially support the needs of a broader range of users.
- The public sector, private sector, non-governmental organizations (NGOs) and academe all have roles in supporting the broader user needs and these roles will change with time.
- All potential user groups represented at the conference believe they would benefit from EOSDIS.
- Within its current resource allocation, EOSDIS cannot support all the needs of its potential user community.
- A variety of current users are supported by existing data information systems. The use of these systems can and will continue, but these operations may be enhanced by EOSDIS.
- Awareness and information about the existence of EOSDIS data, information and services is inadequate.

GENERAL RECOMMENDATIONS

The general recommendations are:

- On a continuing basis, EOSDIS must evaluate users evolving needs and seek adequate feedback.
- EOSDIS should help meet the needs of the broader potential EOSDIS user community directly and through partnerships.
- EOSDIS should continually review its role within the mix of government, university and private sectors in serving user needs.
- EOSDIS should seek ways to enhance its interaction with local, state, national, and international data and information systems, both in the public and private sector.
- EOSDIS must implement a system of governance that includes all segments of its potential user community.
- EOSDIS should seek innovative ways to educate the potential user community about the utility of its data, information, tools and services.
- A follow-on conference of the potential user community should be convened in approximately two years to review the success of EOSDIS in meeting user needs and to provide guidance to t he evolutionary planning of the system.

References

1993 EOS Reference Handbook. Ghassem Asrar and David Jon Dokken, editors. NASA: Washington, DC. August 1993.

John E. Estes
Professor, Department of Geography
Director, Remote Sensing Research Unit
University of California
Santa Barbara, CA 93106-4060
805/893-3649 (phone)
805/893-3703 (fax)
estes@geog.ucsb.edu

http://www.ncgia.ucsb.edu/eosdis/homepage.html for more information on the EOS Potential User Model Development

Potential Commercial Applications Of Earth Observing System Data And Information System Products

Michael Lawless
Joseph Scepan
John E. Estes

Abstract

This preliminary study examines the potential for utilizing science data products of the National Aeronautics and Space Administration Earth Observing System (EOS) in important applications and analysis areas.

Applications potential of each of the science data products from the twenty-one EOS sensors was evaluated. Specifications and parameters for each of the EOS instruments were reviewed and the stated accuracy, temporal resolution, horizontal resolution/coverage, and vertical resolution/coverage of each of their science data products were examined. EOSDIS applications are presented within the context of eighteen individual potential applications areas that were identified to capture the broadest possible range of potential EOS science data product uses.

Results of this analysis are presented as summaries provided for each EOS sensor and potential science data products applications are shown in a matrix format for the most important sensors. Within each matrix data product use is specified in four levels of potential: High (3); Some (2); Possible (1), and; None (0).

Matrices were developed for the following instruments:

Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)
EOS Ocean Color Instrument (EOS-Color)
Earth Observing Scanning Polarimeter(EOSP)
Landsat Enhanced Thematic Mapper Plus (ETM+)
Moderate Resolution Imaging Spectrometer (MODIS)
Tropospheric Emission Spectrometer (TES)

Results of this analysis show that 18 of the 21 total EOSDIS systems are potentially useful in one or more applications areas and 40 separate science data products acquired by these sensors have potential applications.

GIS and Hydrologic Modeling - an Assessment of Progress

by David R. Maidment, Dept of Civil Engineering, University of Texas at Austin

Presented at the Third International Conference on GIS and Environmental Modeling, Santa Fe, New Mexico, Jan 20-25, 1996.

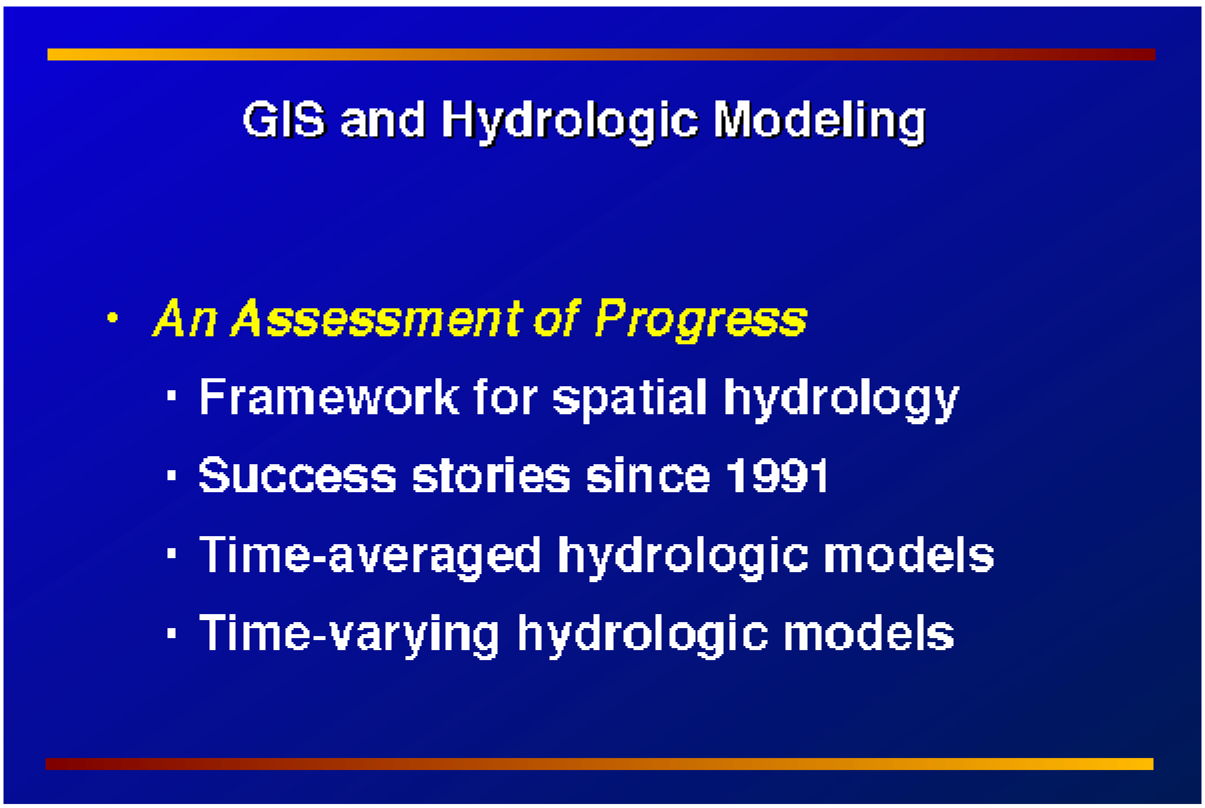


Table of Contents

Introduction

The First International Conference on GIS and Environmental Modeling was held in Boulder, Colorado, in September 1991. At that Conference I presented a survey of the state of GIS and hydrologic modeling as the subject then existed (Maidment, 1993). The intent of this paper is to measure the progress of the subject in the four years that have elapsed since the first conference, to present a framework within which GIS hydrology can be viewed as an integrated subject, and to examine how spatial hydrologic models can be created for time averaged and time varying systems. The focus of this paper is on a personal approach to this field rather than a survey of all the relevant applications.

Important progress has been made during the last four years in the availability of comprehensive spatial data sets which support hydrologic modeling. The distribution of low cost or free data sets via Internet and CD-ROM for digital elevation data, soils, land use, and climate data has stimulated the development of procedures for processing the data into useful forms for hydrology, such as those shown on our GIS hydrology home page (http://www.ce.utexas.edu/prof/maidment/). It has also stimulated the construction of a number of integrated systems where existing hydrologic models have been connected to spatial data bases resident in GIS, with the SWAT (Soil Water Assessment Tool) of the USDA and the MMS (Modular Modeling System) of the USGS being probably the most comprehensive systems of this type.

A further and more complicated question is to ask how hydrologic modeling can be rethought in the spatial context that GIS provides. In other words, instead of attaching existing models to GIS databases, can new hydrologic models be created that take advantage of the spatial data organizing capabilities of GIS? This question implies a reversal of traditional priorities in hydrologic modeling where the emphasis has always been on the way that physical processes are represented, and the manner in which the parameters are to be obtained for a particular environment plays a relatively minor role. In a spatial hydrology model, the emphasis is first on the digital description of the environment, and then on the formulation of process models which can fit the available data and environmental description.

Developing a Spatial Hydrology Model

Hydrology is concerned with study of the motion of the earth's waters through the hydrologic cycle, and the transport of constituents such as sediment and pollutants in the water as it flows. GIS is focused on representing the landscape by means of locationally referenced data describing the character and shape of geographic features. A spatial hydrology model is one which simulates the water flow and transport on a specified region of the earth using GIS data structures. Suppose the boundary of this region is represented by a polygon, such as a river basin boundary or an aquifer boundary. Because both vertical and horizontal water flow can be taking place within the region, hydrologic processes need to be defined over a volume of space rather than an area. Such a volume can be constructed by projecting vertically the lines making up the polygon boundary into the atmosphere and into the earth, and closing the top and bottom of the volume by horizontal planes. Such a volume is called a *control volume* in fluid mechanics and the surface which surrounds it is called the *control surface*.

Building a hydrologic model involves writing equations that relate the rates of change of water properties within the control volume to flow of those properties across the control surface. For example, a simple soil water balance model for a control volume drawn around a block of soil is:

$$S(t+1) = S(t) + P(t) - E(t) - Q(t) \qquad (1)$$

in which S(t) represents the amount of soil moisture stored at the beginning of the time interval t, S(t+1) the storage at the end of that interval, and the flow across the control surface during the interval consists of precipitation P(t), evaporation, E(t), and soil moisture surplus, Q(t), which supplies streamflow and groundwater recharge. Solving this equation requires dealing with time series of the four variables: S, P, E, Q, and possibly of other variables related to them.

It is implicit in constructing a spatial hydrology model that the properties of the system will be spatially variable, so a time series for each of the variables just described must be generated for each soil unit in the domain of analysis. Suppose that there are L spatial units, and that analysis on a single unit requires the definition of M variables for each of N time periods. The number of values to be determined is given by the product LMN, a number that can easily explode beyond the capabilities of available computer memory and reasonable computation times. Indeed, it appears that if the product LMN exceeds a limit of approximately 1 million, solution of the model will be computationally difficult. It follows that in constructing a GIS hydrology model the first task is to determine what variables will be calculated on how many spatial units for a defined number of time periods. Construction of complex models must proceed by partitioning the total problem into a series of submodels that interact with a common database. It is the capacity of GIS for rigorously defining this database which makes possible complex models connecting various parts of the hydrologic cycle within a particular region. One begins with a GIS description of the region and then by modeling adds additional detail to the regional description concerning water flow and constituent transport.

Time and Spatial Domains

In considering these questions, a fundamental distinction arises in the treatment of spatial units depending upon whether raster or vector data structures are used. In a raster data structure, a fine mesh of cells is laid over the landscape and all calculations are done for each cell. This approach, based on using digital elevation model cells as the spatial units, is very useful for certain kinds of hydrologic analysis. But the number of DEM cells within typical analysis regions is usually very large, typically 10,000 to 1 million, so the number of time periods that can analyzed is usually relatively small. Indeed, one may eliminate time as a dimension of the problem by working with mean annual values of all the variables, and the subject of raster based mean annual flow and transport models is presented at further length later in this paper.

If time dynamics are to be considered, it is usually necessary to employ a vector data structure based on related points, lines and polygons. To describe the values on L spatial units of M computed variables in N time periods requires in concept a 3-D data structure but this can be reduced to a set of 2-D data structures in the following manner. The feature attribute table of the GIS coverage defines the geographic properties of the L spatial units and gives each a unique identifying number. This table has geographic attributes as its columns or fields for which the values in each spatial unit are displayed in rows. The values of a particular computed variable, such as soil moisture storage, S, can be defined by means of a related *time table*. The conventional method of constructing such a table is to define a new field for each time interval and keep the rows for the sequence of spatial units, but the advent of object-oriented GIS programming languages, such as the Avenue language in Arcview, has made possible the reverse arrangement, namely the use of time as the index on the rows of the table for which there is a new field for each spatial unit. The item name in this field is the feature identification number of the spatial unit attached to an arbitrary prefix such as SU to mean spatial unit. Thus, in a time table for soil moisture storage, the value of soil moisture storage in time period 154 on spatial unit 30 may be found in row 154 of the table in field SU30.

This arrangement of time vertically and space horizontally makes it easy to see by reading down a column the temporal sequence of values of a variable in a particular spatial unit. The key point in this description is that using object oriented data structures, a value in one table (the feature identification number) can be related to a field in another table instead of to a single value or a set of such values. For hydrologic modeling, the smooth treatment of time variation within GIS is a critical problem that has strongly limited what could be done in the past. It appears that this limitation has now been lifted and the models of spatially distributed and time varying systems can be readily constructed within GIS rather than simply using the GIS as a repository for spatial data feeding an external time varying model.

Ten-Step Modeling Procedure

In Table 1 is presented a 10-step plan by which preparing a GIS hydrology model can be broken down into component parts. The first five of these steps deal with defining the framework of the model in space, and time, and in preparing the environmental description, which may include representation of the land surface terrain, soils and land cover, subsurface hydrogeology, and hydrologic data such as precipitation, streamflow and constituent concentrations. The second five steps deal with simulating the water balance of spatial units, the flow of water and transport of constituents between units, the effect of water utilization structures such as dams and pumping systems, and finally, with the presentation of the study results.

---

1. **Study design:** Objectives and scope of study; spatial and time domain; process models needed, variables to be computed.
2. **Terrain analysis:** Deriving a watershed and stream network layout from digital elevation data and mapped streams.

3. **Land surface:** Describing soils, land cover, land use, cities, and roads.
4. **Subsurface:** Hydrogeologic description of aquifers
5. **Hydrologic data:** Locating point gages, attaching time series and their average values, interpolating point climatic data onto grids.
6. **Soil water balance:** Partitioning precipitation into evaporation, groundwater recharge and surface runoff; partitioning of chemicals applied to the land surface.
7. **Water flow** Movement of water through the landscape in streams and aquifers. Computing streamflow and groundwater flow rates.
8. **Constituent transport:** Transport of sediment and contaminants in water as it flows. Computing concentrations and loadings.
9. **Impact of water utilization:** Locating reservoirs, water withdrawals and discharges in rivers, and aquifer pumping. Their effects on water flow and constitutent transport.
10. **Presentation of results:** Developing visual and tabular presentation of the study results. Use of Internet and CD-ROM to transmit results.

Table 1. A ten-step procedure for a GIS hydrology study.

In our GIS hydrology home page is presented sources of spatial data on Internet that can be used in preparing the environmental description. The advent of STATSGO soil data base and the Internet access to the RF1 river reach file, the USGS Hydrologic Unit Code for watershed boundaries and the USGS Land Use and Land Cover data are all a great help in constructing hydrologic models.

Processing Digital Elevation Data

Probably the principal advance in GIS hydrology modeling that has occurred during the past several years has been the widespread availability of digital elevation data via Internet and CD-ROM and advances in the methods of processing them. DEM data are available in the United States at 30m and 3" cell size, suitable for delineating watersheds and stream networks within urban areas or in small river basins, at 15" cell size which is suitable for regional studies of the size of Texas, to 30" cell size which is appropriate for continental scale studies to 5' cell size covering the earth. The 30" DEM's for Africa and other regions being constructed by Sue Jenson and her co-workers at the USGS using the Digital Chart of the World are a significant contribution.

Delineation of watersheds from DEM data has become standardized on the eight-direction pour point model in which each cell is connected to one of its eight neighbor cells (four on the principal axes, four on the diagonals) according to the direction of steepest descent. Given an elevation grid, a grid of flow directions is constructed, and from this is derived a grid of flow accumulation, counting the number of cells upstream of a given cell. Streams are identified as lines of cells whose flow accumulation exceeds a specified number of cells and thus a specified upstream drainage area. Watersheds are identified as the set of all cells draining through a given cell. The "thousand-million" rule is a rough guide in this activity - take the area of the region to be analyzed and divide it by one million to give the appropriate cell size to use; multiply the cell size chosen by one thousand and that is the minimum drainage area of watersheds that should be delineated from this DEM.

Table 2 shows the cell sizes of digital elevation data and their typical range of application. In this table, a typical watershed area contains 5000 DEM cells and the region of application contains 200 of these watersheds.

Geographic Cell Size	Linear Cell Size	Watershed Area (km2)	Region Area (km2)	Typical Application
1"	30m	5	1000	Urban watersheds
3"	90m	40	8000	Rural watersheds
15"	460m	1000	200,000	River basins
30"	930m	4000	900,000	Nations
3'	5.6 km	150,000	30,000,000	Continents
5'	9.3 km	400,000	90,000,000	Global

Table 2. Digital elevation cell sizes and their scope of application

Standardized Approach to Watershed Delineation

A standardized way of delineating watersheds and stream networks is the following: construct the stream network in the standard way by choosing a drainage area threshold; divide the stream network so created into individual stream links; find the outlet cell at the lower end of each link; delineate the watershed for each of these outlet cells. By changing the threshold drainage area, subwatersheds can be delineated within watersheds in a nicely scaled manner. This algorithm has a critical property - there is one and only one stream for each watershed, which makes modeling of the flow of water from watersheds to the streams within them feasible within GIS. Also, there is a one to one relation between the raster representation of the landscape and the vector features of streams and watersheds derived from it.

Several variations on this standard algorithm are useful. First, the DEM delineated streams are usually close to but do not coincide with mapped streams. Sometimes, critical errors occur when a portion of the upstream part of a basin drains into the wrong downstream river. To overcome these errors, the mapped streams can be converted into a grid and "burned in" to the DEM by artificially raising the elevation of the off-stream cells. This technique requires editing of the stream network to eliminate stray streams and loops that would confuse the delineation process and it results in some distortion of the watershed boundaries in areas where the DEM and mapped streams are not completely consistent, but it has the great advantage that the DEM delineated streams match the mapped streams exactly. After all, it is the stream network which is really the critical item in landscape delineation because it is the stream that carries the water. This "burn in" technique is especially useful in coastal zones with very flat terrain and other locations where drainage is directed through constructed channels. It also helps to ensure that gaging stations and other features are precisely located on the stream. Other variations on the standard delineation technique include the identification of zones of interior drainage and the subdivision of long streams by placing outlet cells along the streams at arbitrarily defined intervals. Arc Macro Language Scripts for watershed delineation are available on our GIS hydrology home page

Time Averaged Hydrologic Modeling

Mean Annual Flow

The determination of flow at ungauged locations is a common problem in hydrology. A simple approach to this problem is to eliminate time as a dimension by restricting the computation to mean annual flows. The analysis can then be constructed by using the cells of a DEM grid as the computational units. One begins with a mean annual precipitation grid over the landscape, which for the United States has been constructed by Daly et al. (1994) and for Africa by Hutchinson et al. (1995), both using approximately 3' cells. The precipitation for each DEM cell is determined from the climate grid. The watersheds of each of the stream gauging stations in the region are delineated and the mean annual precipitation, P, for the drainage areas determined. The longest streamflow record in the basin is used as an anchor record, a long period of analysis is chosen (such as 1961-1990), and the mean annual flow per unit of drainage area, Q, is determined for each gage. If some of the gages have incomplete records, the long term estimate of the mean annual flow can be found by: long term flow at a sample gage = long term flow at the anchor gage x (flow at sample gage / flow at anchor gage) where the ratio in parentheses is constructed using the means of the common period of record at the two gages. A graph is plotted of Q versus P, and an equation fitted to the relation:

$$Q = cP \qquad (2)$$

where c is a runoff coefficient, which is a nonlinear function of the precipitation. In dry areas, the greater is the precipitation, the greater is the percentage of the precipitation which becomes runoff. By multiplying the mean annual precipitation grid by this runoff coefficient, a mean annual runoff per unit area can be determined for each DEM cell. This quantity can be used as a weight and a weighted flow accumulation performed in the same manner as the regular flow accumulation is done when constructing the watershed boundaries. The weighted flow accumulation of each DEM cell, when multiplied by the cell area, gives the mean annual flow for each cell. Thus a mean annual flow map can be derived with estimates of the flow at every stream location in the landscape. This is a very simplified method of hydrologic analysis but one that is faithful to the gauged data in the region and can be applied to large regions in a consistent manner.

Non-Point Source Pollution Assessment

An extension of the mean annual flow estimate just derived can be used to estimate non-point sources of pollution loading to streams. Such sources include pollutants from agricultural areas and from urban runoff from areas such as roads and parking lots. Point pollutant sources are those associated with a particular outlet location, such as the outlet of a wastewater treatment plant. A standard assumption in treating non-point source pollution is to relate the expected concentration of pollutants in runoff to the land use in the drainage basin. By taking a land use map of the area, and using a look-up table to connect land use to expected pollutant concentration, a map of expected concentration, C, can be determined and the expected pollutant concentration in runoff from each DEM cell calculated. By taking the product of the concentration and the flow rate per unit area, a pollutant loading rate per unit area, L, can be calculated for each cell using the relationship:

$$L = QC \qquad (3)$$

Computing the weighted flow accumulation of this loading onto downstream cells and multiplying by the cell area gives the mean annual pollutant loading in each cell. Nice maps of expected pollutant loadings can thus be created. By dividing this mean annual loading by the mean annual flow, a grid of expected pollutant concentration in each cell is derived, which is derived from the weighted average of all the contributions of flow and loading from upstream. Thus is created an expected concentration map, which of course is really meaningful only in the stream cells.

Pollutant concentration is normally sampled at a number of locations in the basin. By making the assumption that the concentration sampled each time at a particular place is drawn from the same water (i.e. that the data are statistically stationary in time), an average observed concentration can be computed for each sample point. The observed average concentrations can be compared with the expected mean annual concentrations computed from the flow accumulation model by drawing a circle at each sample point having an area proportional to the number of samples, and using a consistent color coding scheme for the expected and observed average concentrations. This analysis quickly shows where pollution is at or near expected levels, where point sources are raising the concentration well above the levels expected from non-point sources alone and provides a spatial picture of the variation of observed pollutant levels. Like the mean annual flow analysis, this grid-based non-point source pollution assessment method is simplified but it makes use of the observed hydrologic data and the GIS data normally available in a region in a reasonable way and yields useful results.

Time Varying Water Balance Models

A *water balance model* is a representation of the mass balance of water within a particular control volume. It is a physical statement of the law of conservation of mass which states that matter cannot be created or destroyed. As a result, the rate of change of storage of water within the control volume is equal to the difference between its rates of inflow and outflow across the control surface. One may distinguish in constructing a spatial hydrology model between the surface defining the outer boundary of the study region, and the surface defining the boundary of the spatial units within that region. A spatially distributed water balance model applies the law of conservation of mass to describe the mass balance within each spatial unit, and to this must be coupled a momentum equation (such as Darcy's law for groundwater flow) which defines how quickly water can move between units. Different sizes and shapes of spatial units are needed to deal with the different phases of the hydrologic cycle.

Atmospheric Water

Most water that falls on the land surface is derived from oceanic evaporation carried inland by atmospheric circulation, so it is appropriate to begin the study of hydrology by examining the motion of atmospheric water. The most useful way of doing this in a GIS context is to use the results of GCM modeling, where the acronym GCM means here General Circulation Model (this was the original meaning of this acronym before the more popular Global Climate Model came into vogue). In the United States, the National Meteorological Center in Maryland maintains a global GCM in continuous operation for numerical weather forecasting, which is updated each 12 hours with data from atmospheric soundings obtained from a global network of balloon-borne sensors released from weather stations, called the Global Data Assimilation System. The condition of the atmosphere (temperature, density, wind velocity, air pressure, moisture content) is calculated on a geographic grid of 2 degree cells covering the earth, using a very short time interval of the order of a few minutes, for a time horizon of a few days ahead. Each 12 hours, the forecasts are updated with new observations, and the process repeated. Summary statistics are available from the National Center for Atmospheric Research in Boulder, Colorado. A similar service is provided in Europe by the European Center for Medium Range Weather Forecasting in Reading, England.

An example of the atmospheric moisture flow over North America is provided in our GIS hydrology home page, with atmospheric moisture flow being given by the product of the wind velocity and atmospheric moisture content, vertically integrated over the height of the atmosphere, spatially averaged over the 2 degree grid, and time averaged over a month, for the months June 1991 to August 1993. An animation of these data provided at this home page site, shows that the bulk of the water precipitating over the central and eastern regions of the United States is derived from evaporation in the eastern Caribbean Sea, carried in an enormous wheeling motion over the Gulf of Mexico and the Southern United States, first towards the West and then later towards the East as the moisture moves north of the 30th parallel.

The atmospheric water balance for a region, such as the State of Texas, can be computed by taking the vector field of atmospheric moisture flow, **q**, defined on the 2 degree grid, and interpolating the flow vectors onto points lying on the border of the State at regular intervals. The border of Texas is 7000 km long and points on the border at 100 km intervals seem to be appropriate. The lines joining these border points are also a set of vectors, **L**, which can be imagined to project vertically into the atmosphere and form a closed set of planes surrounding the atmosphere of the State, which are the control surface of an atmospheric control volume drawn over the State. The flow of water across any one of these planes can be found from the vector cross product:

$$Q = q_x L_y - q_y L_x \qquad (4)$$

where the Easting and Northing components of the flow vector **q** = (qx, qy), and the border line vector **L** = (Lx, Ly) are cross multiplied. This is a rather interesting example of two different uses of the concept of a vector: firstly, in the sense of physical science for the atmospheric flow data, **q**; secondly, in the sense of GIS for the border line, **L**. By continuing this computation around the border, the net outflow of atmospheric moisture across the control surface can be determined, which is equal to difference between evaporation and precipitation on the land surface plus the change in atmospheric moisture content within the control volume. It turns out that the difference between atmospheric moisture inflow and outflow over Texas, Q, is usually less than 5% of the average atmospheric moisture flow over the State, so small errors in the flow computation can lead to substantial variations in the value of Q. The net evaporation from the land surface computed from this global data set may not be very accurate. The US National Meteorological Center is presently implementing a new mesoscale GCM over North America called the Eta model, using 40 km computational cells. The results of this model should permit more accurate atmospheric water balance computations to be made for regions like Texas.

Soil Water

Soil water is that water contained within the soil column, so the control volume for the water balance is a block of soil, and the computation consists in relating the change in soil moisture content to evaporation, precipitation, and outflow from the soil. At first sight, it would seem that the most appropriate spatial unit to use for a soil water balance would be a soil map unit, but these map units have very irregular shapes and a great range in size from one map unit to another. Climate data also play an important role and it appears that because the construction of spatially distributed climatic data usually involves interpolating climate onto a grid, that the grid cells used for that interpolation are also appropriate as "soil boxes" for constructing a soil water balance. Such climate cells are approximately 3' in size for the climate grids of the United States constructed by Daly et al. (1994), and for Africa constructed by Hutchinson et al. (1995). Willmott et al. (1985) constructed a global soil water balance model, and later Legates and Willmott (1990) developed a global climatology of monthly mean temperature and precipitation grids on 0.5 degree cells. Willmott uses the simple Thornthwaite soil water balance method, which requires only a single soil water parameter, the water holding capacity. Other soil water balance models, such as that in SWAT, have several soil layers and require more soil parameters. The recent emergence of a satellite derived net radiation balance of the earth (Darnell et al. 1992) provides net radiation estimates for the soil water balance, an important new data source.

The product of a soil water balance is a time history on a daily or a monthly basis of soil moisture content, evaporation and "water surplus" which is the water flowing from the soil to form surface runoff and groundwater recharge. Given the same input data, computation on a daily basis will always yield more water surplus than will computation on a monthly basis because daily precipitation is an episodic process, zero on most days, but when a precipitation event occurs, the soil moisture storage can be quickly filled up, thus producing a water surplus; if the same data are averaged over a monthly interval, it is as if the precipitation falls as a gentle mist, which may evaporate back to the atmosphere before the soil moisture capacity is filled. Interpolation of daily precipitation onto a grid is an uncertain undertaking because the spatial variation in daily precipitation is large. There is thus a challenge in constructing a GIS hydrology model for soil water balance in choosing the appropriate time interval for calculation.

Groundwater

There are two kinds of groundwater flow: *unconfined flow,* which occurs near the land surface and for which the water table or phreatic surface is the upper boundary of the saturated aquifer, and *confined flow* which occurs in deeper aquifers where the upper boundary on the flow is provided by an overlying aquitard, or hydrogeologic unit of low permeability. Unconfined flow is strongly influenced by surface hydrologic conditions, especially by seepage from streams crossing the region where the aquifer outcrops at the land surface. Confined flow is less influenced by surface conditions. Groundwater flow models usually use rectangular or triangular spatial units to represent elementary volumes of porous medium upon which the computations are to be performed. A GIS-based groundwater model can similarly be constructed using polygonal units provided the polygons are reasonably regularly shaped, but groundwater models constructed within GIS are not very computationally efficient.

In constructing a groundwater balance model, there are two computations to be performed: first, a water balance on each spatial unit in which all the inflows and outflows of the unit are used to determine the change in water storage and thus of the piezometric head within the unit; second, a flow computation between each pair of spatial units in which Darcy's law is used to determine the rate of groundwater flow as a function of the difference in head and the flow properties of the aquifer in the units. In a map-based groundwater modeling system, the first computation is done over all the polygons making up the aquifer, while the second is done over all the boundary lines of those polygons. Interaction between surface water in streams and underlying groundwater can be similarly determined by applying Darcy's law to the difference in piezometric head between the stream passing through an aquifer unit and the surrounding aquifer. All these computations need to be done on reasonably small units not more than say 20 km in cell size, because otherwise the head gradients in space become very small. Groundwater aquifers are usually quite confined in area and do not extend over the whole landscape, so unlike surface water flow which takes place everywhere, groundwater flow is more of a localized problem and a regional study needs to take into account each aquifer in the region individually, rather than considering groundwater flow to be a regional phenomenon.

Surface Water

Surface water is water in streams, lakes, wetlands and reservoirs. This water system is in some ways the most complex of all the phases of the hydrologic cycle, because it interacts with the other three phases, namely atmospheric water, soil water and groundwater, because the flow velocity is large compared to the velocity of groundwater flow, and because the flow environment is complicated, depending in part on the characteristics of the land surface and in part on the characteristics of the stream system. Fortunately, this is the area where GIS helps the most because of the detailed description of land surface features which can be presented in GIS. As described earlier, by making a suitable terrain analysis using DEM data, a conceptual model of the surface drainage system can be built up in which each watershed has one and only one stream draining it, and each watershed and stream pair can be assigned the same identification number. The watersheds so constructed are of two types: a source or head watershed in which the stream originates within the watershed, and an intermediate watershed where the stream flows both into and out of the watershed.

The stream network is manipulated so that each stream is represented by a single arc, and the arcs are flow ordered so that the from node is upstream and the to node is down stream. Each stream arc is enclosed within its associated watershed polygon. Watershed boundaries are delineated from each stream junction so at most a node can have two streams flowing into it and one flowing out of it. Three flow variables can be associated with each watershed: "From Flow", "To Flow", and "Polygon Flow". From Flow is that stream discharge at the from node; To Flow is the corresponding discharge at the to node; and Polygon Flow is that discharge which comes into the stream by drainage from the surrounding watershed. Polygon Flow is computed by applying a a unit hydrograph to the water surplus computed by the soil water balance model, and it may also include a component representing exchange of water between the stream and the underlying groundwater aquifer. This implies that the soil water surplus data may have to be spatially transferred from the soil water balance spatial units to the watershed units by using polygon overlay functions. The Polygon Flow can be divided by the length of the stream and added progressively to the discharge along the path of the stream, so in a time-averaged calculation, the discharge at a distance D from the upstream end of a stream of length L is given by:

$$\text{Discharge} = \text{FromFlow} + (\text{PolygonFlow} / L) * D \qquad (5)$$

In time-varying flow, the computation is more complex and stream routing methods such as the Muskingum method (Fread, 1993) are appropriate for computing the time distribution of the To Flow given the time distribution of the From Flow and the Polygon Flow. The time table structure described earlier in the paper is used to record the results of these calculations with a separate table being used for each of the three flow variables, a separate field for each watershed, and time on the vertical axis of the table.

In doing such computations, there is a choice between sequencing the computations "first in space and then in time" or "first in time and then in space", in other words, doing all the computations for a given watershed through time and then moving to the next watershed, or doing the computations for all watersheds for one time interval before moving to the next time interval. For regional hydrologic studies where the watersheds don't influence one another the "first in space then in time" method works best; for more localized hydraulic analyses where water level in the river influences groundwater levels, the simultaneous nature of the interactions makes the "first in time then in space" method necessary. The nature of the process interactions governs the computational sequence.

Water Utilization

All of the preceding discussion is valid for a pristine landscape untouched by human activity. But reservoir construction, pumping of water from rivers and groundwater systems, and discharge of wastewater all have profound effects on surface and groundwater flow and quality. Modeling the effects of water utilization permits GIS hydrology models to be useful as a basis for planning decisions on water facilities. Consider a pumping station withdrawing water from a river. In a spatial sense, this can be represented by a point on a river with attributes describing the time pattern of withdrawals. It is useful to locate this point by its relative distance from the upstream end of the reach (D/L in the nomenclature defined previously). This "proportional aliasing" provides a way of locating objects on river reaches which is somewhat independent of the scale of the map used to define the river reach spatially. River reaches are always longer on maps of larger scale but the relative location of a point remains reasonably stable across scales. Once the withdrawal is located, it can be included in the discharge computation.

Including reservoirs in the computation is more complicated because a reservoir is a fairly complicated system by itself, and its outflow depends on the manner in which the reservoir control works are operated. A standard method of simulating water supply reservoirs is presented by Chow, Maidment and Mays(1988).

Conclusions

In looking back at the paper I prepared for the First International Conference on GIS and Environmental Modeling held in Boulder, Colorado, in September 1991 (Maidment, 1993), and comparing the state of GIS hydrology then to its condition now, the following advances seem the most compelling:

- The ready availability on Internet and CD-ROM of data describing the land surface, especially digital elevation data for land surface terrain, which has made it practical for the first time to delineate watersheds in a few minutes in an automated way, and to compute the hydrologic properties of those watersheds. The ability to build an integrated spatial hydrologic data base for a particular region has greatly improved.
- DEM data can be applied for small scale or large scale problems in a relatively scale-independent manner. The Digital Chart of the World, and the 30" DEMs derived from it by the USGS permit the extension to the world of grid-based hydrologic techniques.
- The interpolation of climatic data, especially mean monthly temperature and precipitation, has advanced to the point where its standardized data sets are very useful for regional hydrologic studies. In effect, the contour lines on the classic climatic atlas have been replaced by fine mesh digital grids of data.
- Comprehensive hydrologic simulation systems using GIS databases are now operational, and being used for analysis of basins of more than 1 million km2 in area, a task that would have been unthinkable a few years ago. Large regional hydrologic models are now within grasp.
- Some progress on integration of processes among different phases of the hydrologic cycle is being accomplished by formulating separate models for each phase and then using GIS spatial data handling capabilities to transfer results from one set of spatial model units to another.
- The advent of object-oriented GIS programming languages has broken the barrier to capturing time variation of spatial processes that was so great a limitation in earlier GIS applications to hydrology.

At the same time, a number of formidable challenges still remain before GIS achieves more of its potential in hydrology:

- The standardized data bases that are generally available are applicable to regional scale analysis of fairly large watersheds. Considerable database development is still needed within cities to support analysis at the scale of small urban watersheds. Probably, 80% of hydrologic modeling is done to solve problems in urban areas, so this limitation is critical.

- Various methods for creating GIS-based models of hydrologic processes are emerging but they have not yet been standardized to the point that they are being applied widely. There is a great need for dispersion of knowledge so that more people can use what is available. Internet will help with this. The emergence of more powerful PC-based GIS systems will also help.
- The integration of hydrologic processes, particularly integration of surface and groundwater flow, is not yet solved very well. Integration of processes across scales of space and time is not well understood. A map can be drawn at any scale, but it is unclear to what extent models can be applied at different scales.
- The impact of water utilization facilities, such as pumping stations and reservoirs, on flow through the landscape, is not well described in spatial hydrology models as yet.
- Subsurface representation of hydrogeologic properties of aquifers is embryonic. There are no standardized databases of hydrogeology, like the ones that exist for soils. A 3-D GIS system has not really emerged yet.
- Water quality modeling in rivers and lakes is sufficiently complex that it is still largely being done in traditional simulation models supported by GIS data. There is not yet much intrinsic water quality modeling within GIS.

But, looking back to the first conference four years ago, there is no question that significant progress has been made in spatial hydrology. Most of the challenges that I cited in my earlier paper are now covered in the accomplishments list of this paper; the set of challenges that I have identified here are largely new ones which hopefully will also largely be addressed during the next few years.

Acknowledgements

The research described in this paper is supported in part by a grant from the Texas Water Resources Institute and the US Geological Survey. The methods described here were developed in cooperation with a team of graduate students at the University of Texas, including Jennifer Benaman, Tom Evans, Ferdinand Hellweger, John McLeod, Pawel Mizgalewicz, Francisco Olivera, Jerome Patoux, Seann Reed, Bill Saunders, Peter Smith, and Ye Zichuan. I am grateful for their insights and assistance.

References

Chow, V.T., Maidment, D.R., and Mays, L.W. (1988) *Applied Hydrology*, McGraw-Hill, New York, pp. 530-537.

Daly, C., Neilson, R.P. and Phillips, D.L. (1994) A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain, *Journal of Applied Meteorology*, 33(2), 140-158.

Darnell, W.L., Staylor, W.F., Gupta, S.K., Ritchey, N.A. and Wilder, A.C. (1992) Seasonal Variation of Surface Radiation Budget Derived from Internatial Satellite Cloud Climatology

Project C1 Data, *Jour. Geophys. Res.*, 97 (D14), 15741-15760.

Fread, D.L. (1993) Flow Routing, in *Handbook of Hydrology*, ed. by D.R. Maidment, McGraw-Hill, New York, pp. 10.5-10.13.

Hutchinson, M.F., Nix, H.A., McMahon, J.P. and Ord, K.D. (1995) *A Topographic and Climatic Data Base for Africa - Version 1*, CD-ROM available from Centre for Resource and Environmental Studies, Australian National University, Canberra ACT 0200, Australia.

Legates, D.R., and Willmott, C.J. (1990) Mean seasonal and spatial variability in gauge-corrected, global precipitation, *Int. J. Climatology*, 10(2), 111-127.

Maidment, D.R.(1993) GIS and Hydrologic Modeling, in *Environmental Modeling with GIS*, ed. by M.F. Goodchild, B.O. Parks, and L.T. Steyaert, Oxford University Press, New York, pp. 147-167.

Willmott, C.J., Rowe, C.M. and Mintz, Y. (1985) Climatology of the terrestrial seasonal water cycle, *Jour. of Climatology*, 5:589-606.

David R. Maidment
Dept of Civil Engineering
University of Texas
Austin Tx 78712
Telephone: (512) 471-0065
Fax: (512) 471-0072
Email: maidment@crwr.utexas.edu

Estimation and Evaluation of Spatially Distributed Model Parameters using the Modular Modeling System (MMS)

G.H. Leavesley, R.J. Viger, S.L. Markstrom, and M.S. Brewer

The spatial and temporal variability of hydrologic, climatic, and ecosystem processes and their interactions require the use of distributed-parameter models that can incorporate this variability. The Modular Modeling System (MMS) is an integrated system of computer software that has been developed to provide the research and operational framework needed to support development, testing, and evaluation of water, energy, and biogeochemical process algorithms and to facilitate integration of user-selected sets of algorithms into an operational model. MMS uses a master library that contains user-developed, compatible modules for simulating these processes and which can be linked using MMS tools to create a model.

A geographic information system (GIS) interface has been developed for MMS to support a variety of GIS tools for use in the (1) delineation, characterization, and parameterization of topographic, hydrologic, and ecosystem features, (2) visualization of spatially and temporally distributed module parameters and variables, and (3) analysis and validation of parameter estimates and model results. The GIS tools have been developed for use with Arc/Info and GRASS. They utilize a variety of digital databases and digital elevation models to provide a number of objective characterization and parameterization techniques for application over a range of spatial and temporal scales. Specific combinations of techniques can be linked with MMS modules to automate the estimation of selected module parameters and write these estimates to a model parameter file prior to module execution. The GIS toolset is demonstrated for a hydrologic and ecosystem management model application on the South Fork McKenzie River, Oregon.

Fiona Ellis

The application of machine learning techniques to erosion modelling

Abstract

Land degradation cannot easily be measured over complex regions using either traditional mathematical erosion models or remote sensing. Another approach is to predict areas prone to land degradation using multi-source datasets to create data driven models. However, while it is generally agreed that the use of multi-source datasets can improve classification accuracy over single source datasets, multi-source datasets can prove problematic when subjected to traditional statistical analysis.

Machine learning techniques such as Decision Tree Analysis (DTA) and Neural Networks (NN) are suitable for the analysis of multi-source datasets, as they are distribution free and non-parametric. In addition, DTA and NN are efficient in terms of data requirements. Whilst these properties make machine learning techniques attractive, it is unclear how to implement these types of inferential models, largely due to uncertainties associated with the effect of training set selection on classification accuracy.

This paper investigates the usefulness of NN and DTA in the modelling of areas prone to land degradation by soil erosion in a study area in NSW, Australia. Specifically the effect of selected sampling strategies and sample size on predictive accuracy of these techniques are investigated.

Introduction

Land degradation is a major problem in Australia, affecting large areas, with degradation of non-arid areas alone estimated at 815000 km^2 (Chartres 1987). Land degradation is the end result of any combination of a large number of factors which damage land, water or vegetation resources and may have many forms, including soil salinity, soil erosion, mass movement and woody shrub infestation (Graham 1992). At present a factor limiting the prevention of land degradation in Australia is insufficient monitoring and assessment of the location of degradation processes (Burch *et al.* 1988).

There are three main approaches which can be followed to assess soil erosion. These are (i) process models used to predict the distribution of eroded land, (ii) mapping through interpretation of remote sensing images, and (iii) ground surveys of erosion.

A recent alternative approach used for spatial analysis of complex relationships, which generally improves classification and prediction accuracy over remote sensing alone, is that based upon the use of multi-source data sets (Srinivasan and Richards 1990). Multi-source data sets are those which come from various sources and sensors. This data can be manipulated in a GIS and Artificial Intelligence (AI) techniques used to analyse and provide predictive capability. AI can be described as a group of computer problem-solving techniques that have been developed to try to imitate decision-making processes or to produce the same results as those processes (Scown 1985). Two techniques that fall within the realm of AI are Neural Networks (NN) and Decision Tree Analysis (DTA).

The aims of this paper are: (1) to determine how Neural Networks and Decision Tree Analysis perform as classifiers of soil erosion, individually and comparatively, and (2) to determine how sampling schemes and sample size used in selection of the training set, affect the accuracy of the resulting classifications.

Study area

The study area, Dicks Creek, is 30 km south east of Yass, New South Wales in the Southern Tablelands of south-eastern Australia (Figure 1). The area is 9317 hectares (93.17 km^2]). This area incorporates the Dicks Creek, Williams Creek and Sawpit Creek sub catchments of the Yass River which flows into the Murrumbidgee River and is a part of the Murray Darling River Basin. The area drains predominantly north to the Yass River from sparsely forested land at elevations of 750 m, down through undulating terrain covered with isolated mature Eucalypts, areas of regeneration and grassland at elevations of 550 m.

The catchments are affected by salinity, rill and gully erosion, and extensive areas of sheet erosion. This study area has been a 'demonstration area' of the Department of Conservation and Land Management (CALM) since 1988, the specific objectives of which are to investigate dryland salinity processes and to demonstrate control methods for salinity and erosion.

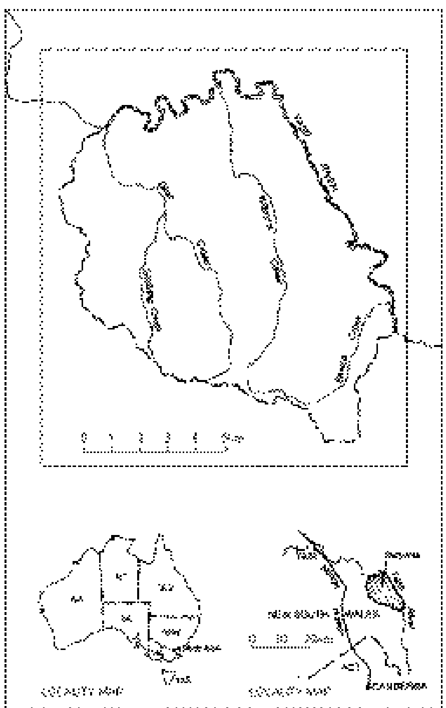


Figure 1 Dicks Creek study area

The area has mainly been used for wool production, with some minor cropping on the river flats, although recent years have seen increasing subdivision of the study area into hobby farms. The dominance of gullying in this region is due to the inherent nature of the soils and changes in the environment, such as land cover changes, since settlement by Europeans in the 1800s (Eyles 1977, Prosser 1991).

Data sources

For the study area (Figure 1), fourteen data layers were compiled and co-registered into a GIS database (Table 1). These variables, thirteen independent and one dependent variable, are easily calculated or derived from cartographic sources and remote sensing. The independent variables include data on physical environmental properties thought to influence erosion distribution, as well as remote sensing data. All data was co-registered to the Australian Map Grid (AMG) and had a grid cell resolution of 30 m, except Landsat TM band 6. The RMS error for geometric correction of the Landsat image was 0.761.

Table 1 Data sources and layers for the Dicks Creek study area

Source	Scale	Data layers derived
Landsat TM 5	-	Bands 1 - 7
Soils Map	1:25000	Soils
Tree Cover Map	1:25000	Tree cover
Topographic Maps	1:25000	Slope
		Aspect
		Flow length
		Accumulated flow
Soil Erosion Map	1:25000	Soil erosion

Landsat Thematic Mapper (TM) bands were used as a surrogate for vegetation cover and hence as a source of information on broad land management practices. The date of the satellite overpass was 13th April 1988, selected to provide a cloud free image, with pixel resolution of 30 m for bands 1-5 and 7, and 120 m for band 6.

The `Soils' and `Tree Cover' maps were digitised from maps produced by CALM (Nicoll and Scown 1993). These maps, along with the `Soil Erosion' map discussed later, were all derived from aerial photograph interpretation and field checking. Colour aerial photographs were used, these being flown in October 1988 at a scale of 1:25000. The soils map contained six soil classes: lithosols, red podzolic soils, soloths, solodics, terrace soils, and alluvial soils. The tree cover map contained ten classes, including categories such as `no mature trees', `plantation' and `regeneration'.

The Murrumbateman 1:25000 Dyeline (CMA 1990) and Bedulluck 1:25000 Topographic sheets (CMA 1980) were used to create a DEM. Contours at 10 m intervals were digitised from these map sheets. The DEM was used to determine slope and aspect through a module of the IDRISI GIS package. The DEM was also used in the derivation of two hydrological indices `accumulated flow' and `flow length'. These four variables are believed to influence the location and severity of soil erosion.

Accumulated flow and flow length were derived from the DEM with Streams/STM4, a local modification of the Jenson (1991) and O'Callaghan and Mark (1984) algorithms. This creates a hydrologically coherent DEM from the topographical map derived DEM. Accumulated flow and flow length, are calculated for each cell. The accumulated flow is equal to the total number of cells that drain into that particular cell, and is an index of soil wetness. The flow length of each cell is calculated as the distance of the path of water from the farthest ridge or high point along the flow direction to that cell, and is a measure of position on slope. Flow length has been used as a surrogate for down slope variation in soil depth, composition and moisture.

Soil erosion is the dependent variable, which is to be predicted from training sets. The map of `soil erosion' that was taken as being `truth' was a soil erosion map prepared by CALM, through aerial photo interpretation and field verification. The aerial photographs used were the same for the `soils' map and `tree cover' map, flown in October 1988 at a scale of 1:25000.

The soil erosion map used in this study contains ten classes, a simplified version of the original soil erosion map compiled by CALM, which contained 20 classes including areas affected by salinisation. Soil erosion in the study area is dominated largely by sheet erosion, this covering approximately 15 percent of the study area (Table 2). Gully erosion covers less than one percent of the study area, occupying spatially discreet locations (Figure 2).

Table 2 Soil erosion classes in the Dicks Creek study area

class	soil erosion class	cells	area (%)
1	no appreciable erosion	86886	83.79

2	minor sheet erosion	15398	14.85
3	moderate sheet erosion	443	0.43
4	severe sheet erosion	37	0.04
5	moderate rill erosion	33	0.03
6	very severe rill erosion	38	0.04
7	minor gully erosion	215	0.21
8	moderate gully erosion	305	0.29
9	severe gully erosion	183	0.18
10	very severe gully erosion	163	0.16

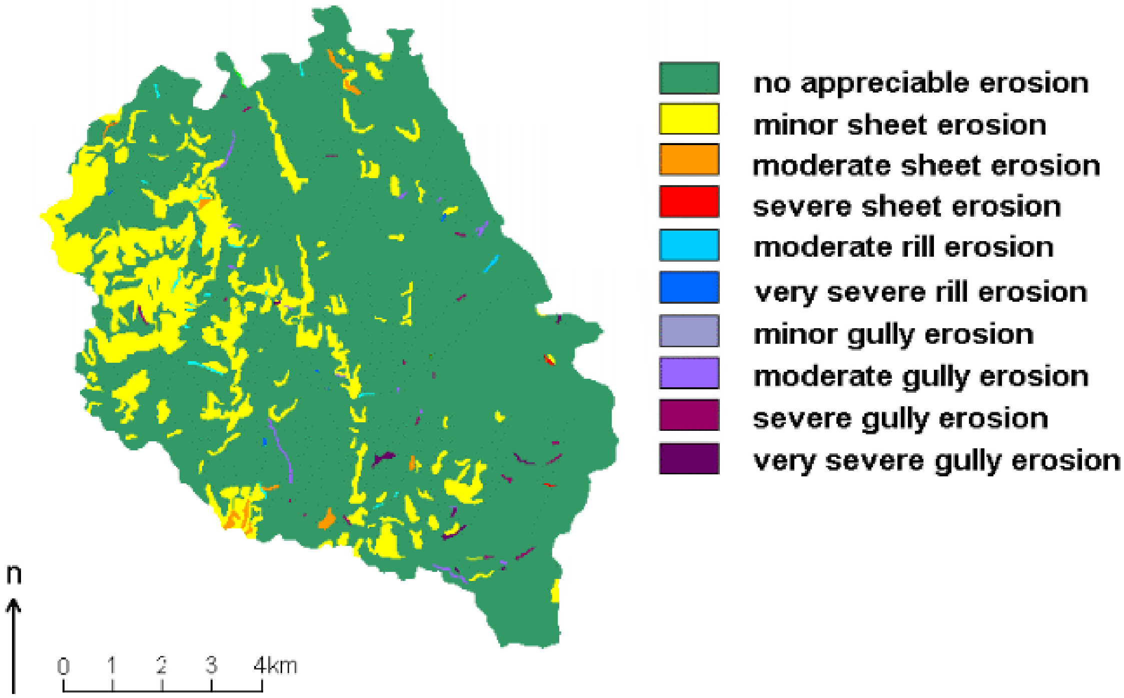


Figure 2 Erosion map of Dicks Creek (after Department of Conservation and Land Management, NSW)

Study framework

Both DTA and NN rely on training sets and it is essential that the training set is statistically representative, reflecting the full range of site diversity. If the training set is not representative of the area, these techniques produce erroneous results.

To investigate this question of sampling strategy and sample size, a simple trial was run. In this trial both DTA and NN were used to classify erosion in the Dicks Creek study area, using training sets of 5%, 20% and 35% of the total number of cells in the study area (103701). That is, training sets had sizes of approximately 5360, 20700 and 36300 cells respectively. These training sets were randomly selected for each repeat.

Standard accuracy assessment methods (error matrices and Kappa) allow comparison of the accuracy of classifications. This is used to determine how sample size affects the accuracy of classifications from DTA and NN in this field area.

The techniques used

Decision Tree Analysis

Decision trees classify data through recursive partitioning of the data set into mutually exclusive subsets which best explain the variation in the dependent variable under observation (Liepins *et al.* 1990, Biggs *et al.* 1991). The decision tree software used here is Knowledge Seeker which is a refinement of the CHAID program used by Kass (1980), allowing the analysis of both continuous and categorical dependent variables. It was chosen because of the partitioning algorithm employed, and because this program allows a high degree of interactive supervision of the 'tree growing' procedure, allowing control of the modelling process if needed. Details can be found in Breiman *et al*. (1984) and Quinlan (1986).

Neural Networks

A two layer `back propagation' neural net was used. Neural networks work on a similar basic principal as DTA, learning from a training data set, storing this relationship, and using these relationships to classify new data, or test data. Details can be found in Benediktsson *et al.* (1990) and Schmuller (1990). The network consisted of 13 input nodes (the independent variables), 6 hidden nodes and 11 output nodes (10 erosion classes and one to mask the study area).

A total of 90 decision trees and 90 neural nets were constructed, 30 replicates of 5%, 20% and 35% samples.

Results

Decision Tree Analysis

The trained decision trees were used to classify the entire Dicks Creek study area. The rules were used to reclassify the thirteen data layers to predict the location and intensity of erosion. One of the resulting images from DTA is shown in Figure 3.

Some pixels were not classified into any erosion class. This because the training area does not always describe all elements of the landscape present across the entire study area. In such cases the decision tree is unable to produce an estimate of erosion for cells with these values.

In order to assess the accuracy of the thematic maps produced, cross correlation matrices were established (eg.Table 3). These matrices express the proportion of each class assigned to a particular category, relative to the actual category as verified through ground observations (Story and Congalton 1986). The columns represent the `truth' (soil erosion map produced through ground observations by CALM) and the rows represent the `predicted' images.

In the rows, a class of zero occurs because the DTA was trained on environmental variables, whose range did not explain the full environmental range present within the study site. In classifying the entire area some pixels were classified as `zero' as they did not fit into any of the erosion classes from the information the DTA was supplied with.

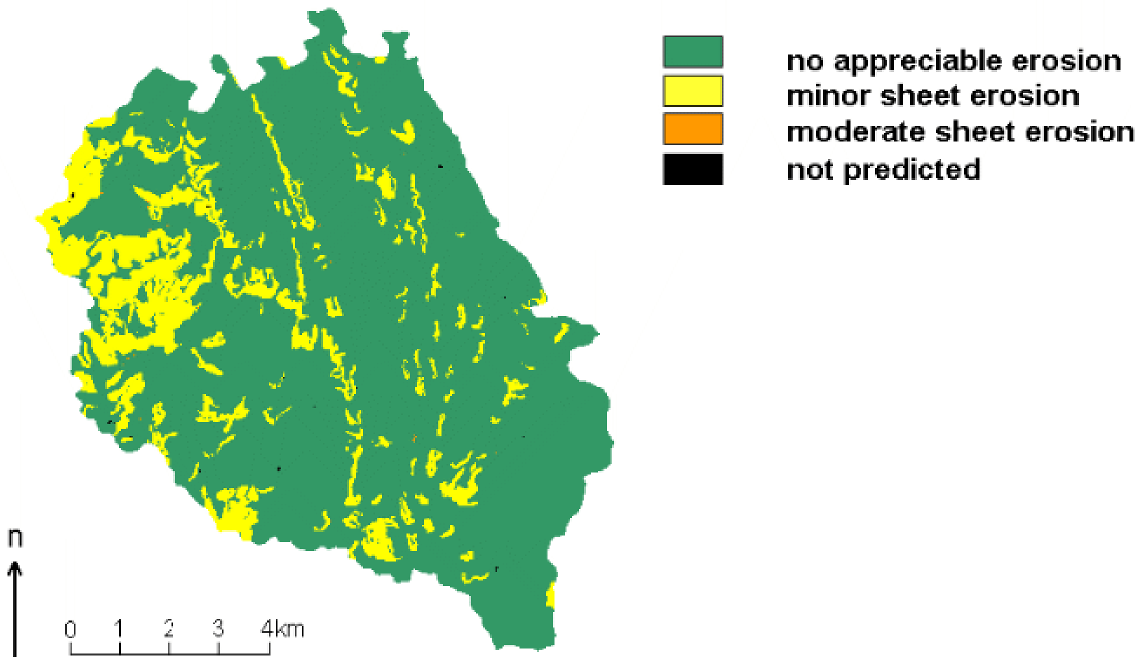


Figure 3 Dicks Creek erosion as predicted by Decision Tree Analysis with 5% training set

Table 3 Cross correlation matrix for Dicks Creek - Decision Tree Analysis classification, using 5% randomly selected training area (a1_5ci.img - overall accuracy 91.92%, overall kappa 0.7017)

	truth										
predicted	1	2	3	4	5	6	7	8	9	10	total
0	31	11	0	1	0	0	0	0	0	0	43
1	83150	3210	86	28	33	38	176	302	181	163	87367
2	3671	12171	356	8	0	0	38	2	1	0	16247
3	34	6	1	0	0	0	1	1	1	0	44
4-10	0	0	0	0	0	0	0	0	0	0	0
total	86886	15398	443	37	33	38	215	305	183	163	103701

Table 4 Cross correlation matrix for Dicks Creek - Decision Tree Analysis classification, using 35% randomly selected training area (1g_35ci.img - overall accuracy 90.93%, overall kappa 0.6673)

	truth										
predicted	1	2	3	4	5	6	7	8	9	10	total
0	1202	36	2	1	0	0	3	0	0	2	1246
1	82672	3737	96	27	31	38	161	297	175	149	87383
2	2888	11564	326	9	0	0	41	1	1	0	14830
3	29	45	18	0	0	0	1	0	0	0	93
4	0	0	0	0	0	0	0	0	0	0	0
5	13	3	0	0	2	0	0	0	0	0	18
6	0	0	0	0	0	0	0	0	0	0	0
7	15	6	1	0	0	0	9	0	0	0	31
8	8	2	0	0	0	0	0	6	0	0	16
9	29	2	0	0	0	0	0	1	7	0	39
10	30	3	0	0	0	0	0	0	0	12	45
total	86886	15398	443	37	33	38	215	305	183	163	103701

Neural Networks

As for the DTA, the trained neural networks were used to classify the Dicks Creek study area, the NN being used to reclassify the thirteen independent variables to predict the spatial location and class of erosion. One output image is shown in Figure 4. Cross correlation matrices were constructed for all replicates. Two matrices are shown (Table 5 and Table 6).

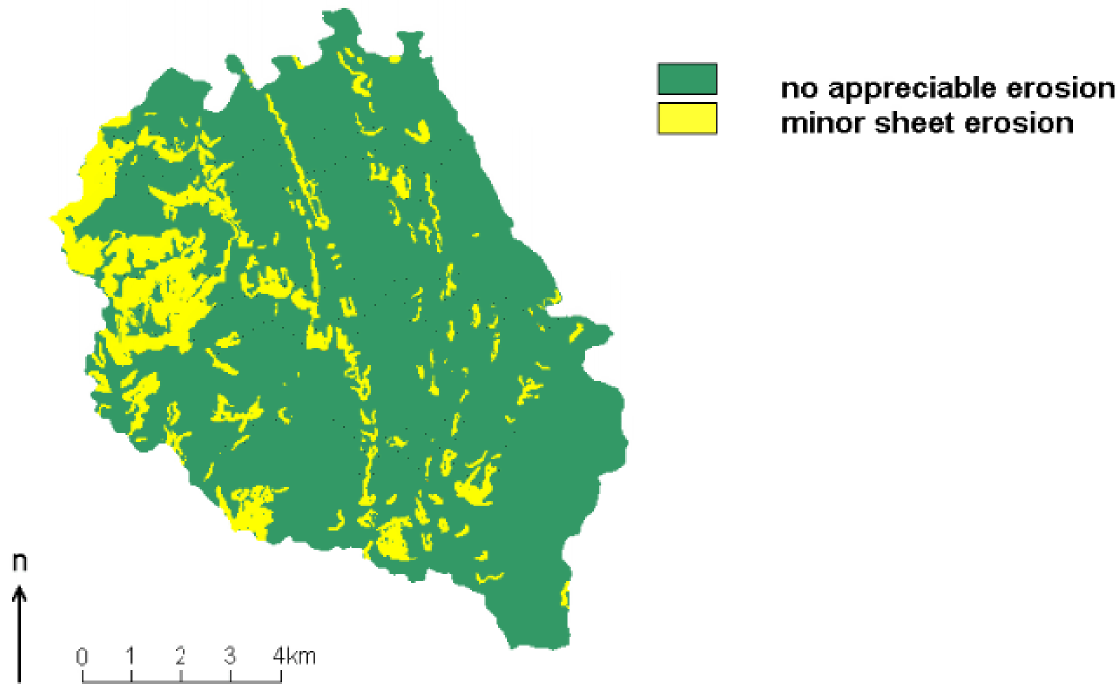


Figure 4 Dicks Creek erosion as predicted by Neural Networks with 5% training set (n1_5ib.img)

Table 5 Cross correlation matrix for Dicks Creek - Neural Network classification, using 5% randomly selected training area (n1_5ib.img - overall accuracy 91.94%, overall kappa 0.7013)

	truth										
predicted	1	2	3	4	5	6	7	8	9	10	total
1	83201	3262	89	32	33	38	168	303	182	163	87471
2	3685	12136	354	5	0	0	47	2	1	0	16230
3-10	0	0	0	0	0	0	0	0	0	0	0
total	86886	15398	443	37	33	38	215	305	183	163	103701

Table 6 Cross correlation matrix for Dicks Creek - Neural Network classification, using 35% randomly selected training area (n1_35ib.img - overall accuracy 91.96%, overall kappa 0.7007)

	truth										
predicted	1	2	3	4	5	6	7	8	9	10	total
1	83315	3351	93	34	33	38	169	303	182	163	87681
2	3571	12047	350	3	0	0	46	2	1	0	16020
3-10	0	0	0	0	0	0	0	0	0	0	0
total	86886	15398	443	37	33	38	215	305	183	163	103701

Table 7 Summary of results

	DTA			NN		
	5%	20%	35%	5%	20%	35%
mean overall accuracy (%)	91.2	91.7	91.9	91.8	92.0	91.9
mean overall kappa	0.67	0.69	0.70	0.69	0.69	0.69
average number of classes predicted	4.5	7	8.3	2	2	2

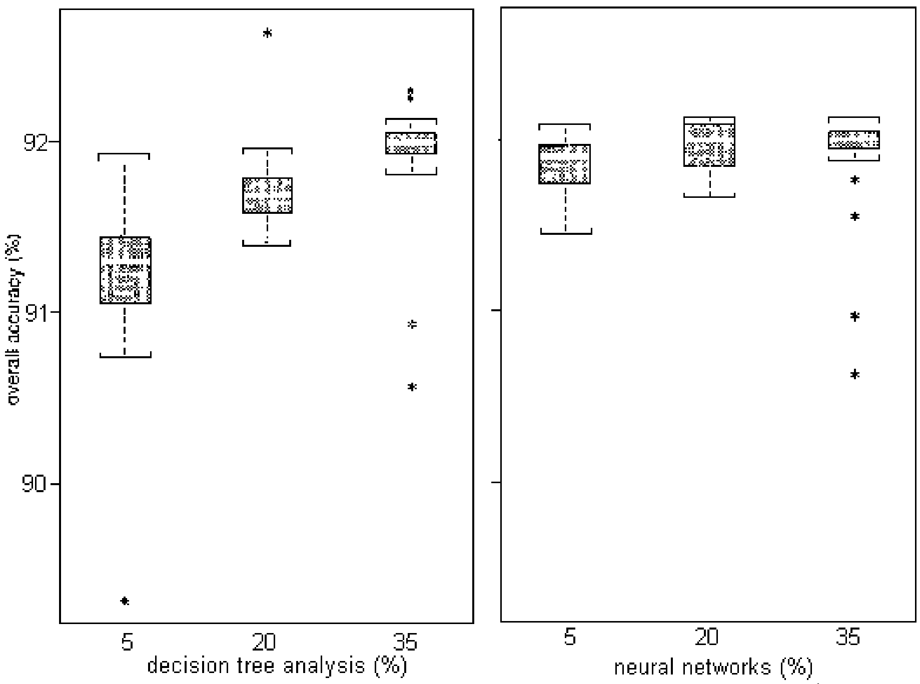


Figure 5 Box plots of overall accuracy for Decision Tree Analysis (5%, 20%, 35% training set size) and Neural Networks (5%, 20%, 35% training set size)

Discussion

From the correlation matrices (Tables 3-6) and from Table 7, it was found that the overall accuracy of the predicted classification was excellent for DTA and NN for all sample sizes, with accuracies of between 91 and 92%.

The results shown in Figure 5, box-plots of the 30 repeats of each sample size, indicate that there is some change in the spread of the data, ie for 35% the overall accuracy and overall kappa values are more tightly clustered around the mean than the 5% values. However, there is very little difference between the accuracies achieved between DTA and NN, both between these two AI classifiers and within them.

An interesting problem is highlighted by this data set. In this study area two of the ten erosion classes, namely `no appreciable erosion' (class 1) and `minor sheet erosion' (class 2) represent 99% of the entire study area, so the training set is swamped by these extensive classes. The remaining 8 classes of erosion (classes 3-10), which include the more severe classes of sheet and gully erosion, occupy extremely small, spatially discreet areas within the landscape. Thus these classes cannot be sampled for adequate training of the AI methods using random sampling.

The high classification accuracies obtained by the AI methods are misleading in terms of the number of classes correctly predicted. In this study, the small classes were often not predicted by DTA and NN, being misclassified mostly into the class of `no appreciable erosion' and, less frequently, into the class of `minor sheet erosion'. The classification of areas of severe soil erosion into the class of `no appreciable erosion' has serious consequences if these methods were to be employed in the field. With training set sizes of 20% and 35%, DTA did correctly predict some cells of these small erosion types (see Table 4).For this particular dataset, some form of stratified sampling would best be employed to enable the prediction of the under-represented classes.

DTA on average identified more erosion classes than NN. This can be explained by the fact that the DTA were run exhaustively, that is until no more splits could be found. If the trees were `pruned' back and then tested, fewer classes may have resulted.

Conclusion

Both DTA and NN achieved high numerical accuracy, when considering the overall accuracy and the Kappa statistic. However, in terms of their predictive accuracy for modelling soil erosion, at this scale and in this study area, they performed poorly. These techniques, when using simple random sampling, could not correctly predict many of the small erosion classes. No matter what the sample size, there is little difference between the results from the two AI classifiers, both within classifiers and between classifiers.

In contrast to many currently available process-based erosion models, these AI techniques successfully use minimal, low cost data sets. DTA allows, through the production of a rule set, a model which is easy to interpret and leads to hypotheses that may explain observed erosion phenomena and guide further investigation.

Acknowledgments

I would like to thank Jeanne di Meglio for her assistance, Brian Lees, Ken Johnson and Richard Aspinall for comments on the draft.

References

Benediktsson, J.A., Swain, P.H. and Ersoy, O.K. (1990) Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing* 28(4): 540-552.

Biggs, D., de Ville, B. and Suen, E. (1991) A method of choosing multiway partitions for classification and decision tree. *Journal of Applied Statistics* 18(1): 49-62.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and Regression Trees*. Belmont, California: Wadsworth International.

Burch, G., Graetz, D. and Noble, I. (1988) Biological and physical phenomena in land degradation. In *Land Degradation: Problems and Policies* (Chisholm, A. and Dumsday, R., eds.), pp. 27-48. Cambridge: Cambridge University Press.

Chartres. C. (1987) Au stralia's land resources at risk. In *Land Degradation: Problems and Policies* (Chisholm, A. and Dumsday, R., eds.), pp. 7-26. Cambridge: Cambridge University Press.

CMA (1980) *Topographic Map 1:25000 Bedulluck (8727-IV-N).* Bathurst, New South Wales: Central Mapping Authority.

CMA (1990) *Topographic Map 1:25000 Murrumbateman Dyeline*. Bathurst, New South Wales: Central Mapping Authority.

Eastman, J.R. (1992) *IDRISI Version 4.0 Technical Reference*. USA: Clark University.

Eyles, R.J. (1977) Changes in drainage networks since 1820, Southern Tablelands, N.S.W. *Australian Geographer* 13: 377-386.

Graham, O.P. (1992) Survey of land degradation in New South Wales, Australia. *Environmental*

Management 16(2): 205-223.

Jenson, S.K. (1991) Applications of hydrologic information automatically extracted from digital elevation models. *Hydrological Process* 5: 31-44.

Johnston, R.M. and Barson, M.M. (1990) *An Assessment of the Use of Remote Sensing Techniques in Land Degradation Studies*. Bureau of Rural Resources Bulletin No. 5. Canberra: Australian Government Publishing Service.

Kass, G.V. (1980) An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* 29(2): 119-127.

Lees, B.G. and Ritman, K. (1991) Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environmental Management* 15(6): 823-831.

Liepins, G., Goeltz, R. and Rush, R. (1990) Machine learning techniques for natural resource data analysis. *AI Applications* 4(3): 9-18.

Neil, D. and Fogarty, P. (1991) Land use and sediment yield of the Southern Tablelands of New South Wales. *Australian Journal of Soil and Water Conservation* 4(2): 33-39.

Nicoll, C. and Scown, J. (1993) *Dryland Salinity in the Yass Valley: Processes and Management*. Final Technical Report for The Natural Resources Management Strategy, Department of Conservation and Land Management. Canberra: Australian Government Publishing Service.

O'Callaghan, J.F. and Mark, D.M. (1984) The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics and Image Processing* 28: 323-344.

Prosser, I.P. (1991) A comparison of past and present episodes of gully erosion at Wangrah Creek, Southern Tablelands, New South Wales. *Australian Geographical Studies* 29(1): 139-154.

Quinlan, J.R. (1986) Induction of Decision Trees. *Machine Learning* 1: 81-106.

Scown, S.J. (1985) *The Artificial Intelligence Experience: an introduction*. Massachusetts: Digital.

Schmuller, J. (1990) Neural networks and environmental applications. In *Expert Systems for Environmental Applications* (Hushon, J.M., ed.), pp. 52-68. Washington D.C: American Chemical Society.

Srinivasan, A. and Richards, J.A. (1990) Knowledge-based techniques for multi-source classification. *International Journal of Remote Sensing* 11(3): 505-525.

Story, M. and Congalton, R.G. (1986) Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing* 52(3): 397-399.

Fiona Ellis
PhD Student
Department of Geography
School of Resource and Environmental Management
Faculty of Science
Australian National University
Canberra ACT 0200
AUSTRALIA
phone +61 6 2798148
fax +61 6 2493770
email fiona@dart.anu.edu.au

D.-I. Müller-Wohlfeil, W. Lahmer, V. Krysanova, A. Becker

Topography-based hydrological modelling in the Elbe drainage basin

The paper is focused on the applicability assessment of two models for large scale hydrological simulation, both of which are based on the topographic index concept. With the first one, TOPMODEL, scaling studies were performed in a number of catchments of different size. Subsequently, an integrated approach coupling the other simplified hydrological model WET with a GIS was applied to the Elbe drainage basin to delineate those areas that are vulnerable as regards water availability. The potential use of TOPMODEL for large scale applications was investigated by studying the effects of different spatial resolutions, area size and topographic conditions on the simulation results. It was shown that runoff estimations are scale-dependent. However, differences in discharge can be compensated by changing the mean transmissivities. Based on these results from the scaling studies with TOPMODEL, application of WET was justified for larger scales. After some modifications with respect to distributed soil and climate data, WET was applied for the Elbe basin. The analysis of water availability was performed in three steps: 1) calculation of the long-term average monthly soil moisture index and evapotranspiration, 2) delineation of subwatersheds in the Elbe basin, 3) identification of critical subbasins from the distribution of the soil moisture index in summer months. The method allows delineation of critical subareas within the study region and can be used for preassessment in large scale hydrological and water quality studies.

---

1. Introduction

1.1 Background

Human induced global climate and land use changes directly affect water and biogeochemical cycles, vegetation structure and plant productivity. However, spatial and temporal patterns of changes may differ in different regions, like the patterns of temperature and precipitation derived from simulation runs performed with general circulation models (GCMs). New methods must be developed to predict regional and local changes in terrestrial ecosystems, as the control measures may be regionally specific. On the other hand, it is necessary to provide inputs and feedback mechanisms to global models of climate and biogeochemistry.

Topography-based models like TOPMODEL (Beven and Kirkby, 1979), TOPOG (O'Loughlin, 1981) and WET (Moore et al., 1993) provide a simple way to introduce lateral flow components into regional or global ecosystem models. TOPMODEL and WET are based on the assumption that local soil moisture dynamics strongly depends on the size of the upslope area (*a*) drained through an observed catchment point, the local surface topographic slope (*tan ß*) representing the hydraulic gradient for saturated water flow, and the downslope soil transmissivity (*T*).

In the first part of this study the suitability and shortcomings of TOPMODEL in large scale hydrological studies are investigated. The model was applied to various smaller catchments and to the German part of the Elbe drainage basin (96,000 km²) as a whole. In the second part of our investigation an integrated approach coupling the simplified hydrological model WET (based on a concept similar to that of TOPMODEL) with a GIS was applied to the Elbe drainage basin to delineate vulnerable subareas.

1.2 Study areas

The Elbe drainage basin is one of the largest river basins in Western Europe. The elevation varies from sea level in the Pleistocene slightly hilly lowlands up to 1161 m in the mountainous areas in Saxony. A significant proportion of sandy soils with high infiltration, low amounts of precipitation, and a high water demand (both climatic and anthropogenic) characterize the high hydrological vulnerability (Becker et al., 1995).

Two smaller basins (Stör in the Elbe basin and Vils in southern Germany) and their subcatchments were chosen for scaling studies. The upper Stör subbasin (1153 km²) is dominated by older glacial and glacifluvial sediments. Dominant soil texture classes are sands with different percentages of loam. More detailed simulations were performed for the Buckener Au subcatchment (58 km²), which is dominated by cropland and forest areas. The maximum elevation is about 93 m a.s.l..

The Vils catchment (756 km²) is part of a dry hilly upland region with elevations between 356 and 613 m a.s.l. Soil texture is sandy to loamy and the vegetation cover is dominated by pine forest and agriculture plants. The Frankenohe subcatchment (45.1 km²), studied separately, is located in the northern part of the Vils catchment.

The simulations in the Stör and Vils basins were based on DEM data sets with a spatial resolution of 1 arcsecond (equivalent to a grid size of about 33 m x 33 m and 29 m x 29 m, respectively) provided by the "Amt für Militärisches Geowesen" (Federal Armed Forces Geographic Office, Euskirchen, Germany FAFGO). For the simulation runs in the Elbe basin a 30 x 50 arcseconds (~ 1000 m x 1000 m) DEM of the "Institut für Angewandte Geodäsie" (IFAG, Frankfurt, Germany) was used.

2. Methodology

2.1 Objectives

Main goals of the study were

- to investigate to what extent low resolution DEMs can be used for topography-based hydrological and water balance studies and to determine potential errors due to the neglect of detailed relief structures,
- to study the influence of relief type (i.e. lowland/mountainous areas) on the simulation results and to investigate possible limitations in model applications,
- to test the applicability of topography-based models for the identification of vulnerable subareas in a large drainage basin with respect to water availability.

In addition, we intended to check whether simplified models (including only some key factors of system behaviour) can be effective tools for the analysis and preassessment of the hydrological cycle in watersheds.

In order to investigate potential limitations in regional applications, simulations based on high resolution DEMs (1 arcsecond) were commenced, which were subsequently aggregated to low resolution DEMs (up to 1 km mesh size). Following Beven (1995), the effects of subgrid heterogeneity are expected to be more important with increasing scale and mesh size. Other physiogeographical factors like geology, vegetation, and land use are not considered. The only additional spatially distributed information are soil data in the case of the WET model applications.

Critical areas for water availability can be defined in a number of ways. In general, the dynamics of soil moisture patterns reflect the overall water balance and can be considered the most important variable defining water availability for vegetation. In order to define critical areas we estimated the long-term average monthly soil moisture dynamics as a component of the water balance, based on long-term average climatic data, topography, and soil. After averaging the soil moisture distribution for subareas of interest (larger grid cells, subbasins or administrative subunits), vulnerable subregions can be delineated.

2.2 TOPMODEL

TOPMODEL (Beven and Kirkby, 1979) is best suited for small to medium catchments (500 km²) with shallow soils and moderate topography which do not suffer from excessively long dry periods. Input data are DEMs and time series data for precipitation and potential evapotranspiration. Data for measured discharges can be used for validation. Output data include simulated discharges, actual evapotranspiration, and information on the build-up of soil moisture and averaged soil moisture deficit.

The *ln(a/tanß)* values were calculated from digital elevation data for every single grid cell of the catchment using a modified version of the GRIDATB programme (kindly provided by Dr. Quinn/ADAS, Wolverhampton).The algorithms of GRIDATB allow multidirectional flow in 8 directions. In order to suppress the generation of very large *ln(a/tanß)* values, the upslope drainage area for each cell can be limited by a threshold (Channel Initiation Threshold CIT) which supports the creation of a virtual river net (Quinn et al., 1995b).

In order to ensure potential flow from each grid cell to its neighbours, the raw elevation data had to be cleaned (i.e. sinks and plateaus had to be removed) before creation of the *ln(a/tanß)* distributions. A modified version of the program SINKS (also provided by Dr. Quinn) was used for this purpose.

Climatic input were daily data of precipitation and potential evapotranspiration, calculated by the algorithm of Priestley and Taylor (1972) and by a method using only daily values of air temperature and potential sun shine hours (Leavesley et al., 1983) respectively.

2.3 WET

The WET model developed at the Australian National University (Moore et al., 1993) is based on a topographic index very similar to that of TOPMODEL. It allows the estimation of the spatial distribution of the long-term average soil moisture and evaporation using an equilibrium approach. Originally the WET model was applied to a small forested catchment (27 km ²) in Australia to estimate spatial patterns of average annual soil moisture and evapotranspiration. The simulations were based on a 30 x 30 m DEM, using a single precipitation value (average annual) for the whole catchment, and without accounting for variation in transmissivity. The annual and monthly time steps are suggested by the authors for WET application (Moore et al., 1993).

The potential evapotranspiration is estimated from the Priestley & Taylor equation for well-watered vegetation under conditions of minimal advection. The equations for the wetness index, potential and actual evapotranspiration are solved iteratively using the Newton-Raphson method, beginning with the element of highest elevation and finishing with the element of lowest elevation at the catchment outlet. WET uses some outputs of the TAPESG (accumulation areas) and SRAD (net radiation) submodels included in the TAPES-G package (Moore et al., 1993). The specific catchment area is estimated by a quasi-random "Rho8"-algorithm that permits drainage from a node to multiple nearest neighbors on a slope-weighted basis. The Geographical Information System GRASS (U.S.Army, 1987) was additionally used for spatial analysis of map layers.

The delineation of vulnerable areas was performed for the Elbe drainage basin in three steps:

1. Calculation of the long-term average monthly soil moisture index and evapotranspiration using WET with distributed soil (water holding capacity) and climate (precipitation and temperature) data for 1 km resolution.
2. Delineation of subwatersheds in the Elbe drainage basin with reasonable average area using the *r.watershed* function of GRASS.
3. Estimation of the average soil moisture index in summer months for subwatersheds, and delineation of critical subwatersheds.

3. Scaling studies with TOPMODEL

Scaling studies to identify potential sources and the magnitude of errors in low resolution topography-based applications have mainly been performed with TOPMODEL. In most of these studies the grid size varied between 2 m and 480 m (Franchini et al., 1996; Bruneau et al., 1995; Quinn et al., 1995a; Wolock and McCabe, 1995; Wolock and Price, 1994; Zhang and Montgomery, 1994; Charait and Delleur, 1993).

Other studies have used components of the TOPMODEL concept with high resolution Digital Elevation Models (DEM) for multiscale or macroscale modelling of hydrological processes and soil-vegetation-atmosphere interaction (e.g. Famiglietti and Wood, 1994; Quinn et al., 1995b).

The main reasons why approaches based on the topographic index (particularly TOPMODEL and WET) have rarely been applied to large areas are that:

- the models were developed for the hillslope/catchment scale and not for the drainage basin scale,
- slope lengths are usually small, and small cell sizes therefore are preferred in TOPMODEL applications (Quinn et al, 1995a),
- simulations can be limited to representative subcatchments, if the region is hydrologically similar in nature (Beven and Quinn,1994).

Since regional modelling approaches have to cope with that data which is available it was decided to apply TOPMODEL at the regional scale with low spatial and temporal resolution. Currently, the spatial resolution of DEM data available world-wide is not better than 1 km (Arnell, 1993) and the temporal resolution of climate data measured operationally is rarely better than 1 day. On the other hand, even for low resolution data, the average location of the upper groundwater table may be well correlated to topography. Smaller scale topography-induced dynamics may have effects on the overall catchment behaviour even in flatter areas and topographic influences cannot be neglected in regional studies. Phenomena like subsurface stormflow and saturation excess overland flow need not be locally limited but may exceed areas larger than the grid size of, e.g., 1 km^2 (Blöschl and Sivapalan, 1995). Their effects should be maintained within the model at this scale.

3.1 Analysis of the Topographic Index

The influence of spatial resolution and catchment size on the *ln(a/tanß)* distributions and the simulation results were analysed in detail for the Stör and Vils catchments. The simulation runs performed for the Elbe basin were focused on studies of the topographic index in order to extend this analysis to a whole landscape. However, no calibration was possible due to missing measured discharge time series data.

In order to study spatial resolution effects, small time period simulations were performed at different resolutions for the Buckener Au subcatchment (32 to 250 m), the Upper Stör catchment (250 to 1000 m) and the Vils catchment (50 to 1000 m). DEM aggregation was performed by the GRID resampling procedure of the Geographic Information System ARC/INFO (ESRI,1991) using the nearest neighbour relationship.

The *ln(a/tanß)* distribution functions are influenced by the spatial resolution as follows:

- Since the number of samples contributing to the *ln(a/tanß)* distributions decreases with decreasing resolution, distributions of high resolution DEMs are smoother than those of low DEMs for the same area.

- Minimum, mean, and maximum of the *ln(a/tanß)* distributions increase with decreasing spatial resolution, due to increasing grid cell area and decreasing slope related to smoothing effects (smaller *tanß* values).

Table 1 summarizes the statistics of the *ln(a/tanß)* distributions obtained for all catchments and spatial resolutions.

				ln(a/tanb)	
catchment	**area [km²]**	**spatial resolution [m]**	**CIT [km²]**	**range**	**mean**
Buckener Au	58	33	0.1	3.26 -14.75	8.94
		110	2	5.80 - 17.74	10.41
		250	4	7.05 - 18.29	11.22
Stör	1780	250	4	6.91 - 18.46	12.56
		500	8	8.550 - 19.20	13.08
		1000	30	9.53 - 20.68	13.77
Frankenohe	45	29	0.03	4.16 - 13.57	8.01
		110	1	7.35 - 13.76	9.97
		250	2	7.35 - 14.64	10.16
Vils	756	50	1	3.40 - 16.60	9.23
		100	2	4.40 - 18.00	9.78
		250	4	6.40 - 18.80	10.78
		500	8	7.40 - 19.80	11.85
		1000	30	8.40 - 20.80	12.67

Table 1: Statistics of the spatial resolutions and the ln(a/tanb) distributions for all (sub)catchments used for the scaling studies

For the same spatial resolution (250 m) and the same CIT value (4 km²), the *ln(a/tanß)* distributions obtained for the four subcatchments studied in detail do not show a general trend, though one could presume a larger range of potential *ln(a/tanß)* combinations for larger catchments (more complex landscapes).

The effects of topography and area size on the *ln(a/tanß)* distributions are most evident for the rather different subcatchments of the Elbe basin. For the German part of the Elbe basin 57 subcatchments were derived using the *r.watershed* function of GRASS (mean area 1762 km²). In general, the mean *ln(a/tanß)* values are lower in mountainous and small subcatchments (smaller *a* and higher *tanß* values) than in flat and large catchments. The mean *ln(a/tanß)* values varied between 11.73 in the southern mountainous regions and 14.94 in extremely flat areas (i.e. the Elbe outlet).

3.2 Scaling Effects

Spatial resolution and area size cause changes in the predicted runoff. The simulation results confirm conclusions drawn by Wolock and Price (1994) according to which the ratio of overland flow to total flow increases with decreasing spatial resolution. Simultaneously, the efficiency (defined as ratio of the variance of the residuals to the variance of the observed discharges according to Nash and Sutcliffe, 1979) is reduced by up to about 41% (see **Table 2**).

			a) identical parameter sets			b) calibrated parameter sets		
catchment	**resolution [m]**	**mean ln(a/tanb)**	**ln(T0)**	**eff-diff [%]**	**diff Qb/Qtot [%]**	**ln(T0)**	**eff-diff [%]**	**diff Qb/Qtot [%]**
Buckener Au	33	8.94	1.72	-20.1	+21.4	0.25	-0.1	+0.7
	110	10.41	1.72	0	0	1.72	0	0
	250	11.22	1.72	-5.9	-13.4	2.53	+0.5	-0.3

Stör	250	12.56	4.36	-1.3	+1.1	3.84	-7.2	-5.7
	500	13.08	4.36	0	0	4.36	0	0
	1000	13.77	4.36	-15.1	-10.2	5.05	-0.1	+3.0
Frankenohe	29	8.01	0.32	-22.4	+31.6	-1.64	+0.2	+0.4
	110	9.97	0.32	0	0	0.32	0	0
	250	10.16	0.32	-5.6	-4.4	0.51	-3.1	-0.6
Vils	50	9.23	0.20	-15.6	+26.7	-1.35	-0.4	-2.0
	100	9.78	0.20	-8.5	+18.6	-0.80	+0.6	-0.2
	250	10.78	0.20	0	0	0.20	0	0
	500	11.85	0.20	-10.8	-23.9	1.27	-1.4	-2.7
	1000	12.67	0.20	-34.7	-40.8	2.09	-1.8	-4.5

Table 2: Influence of T0-shift on the TOPMODEL simulation results
Changes of efficiencies (eff-diff) and base flow to total flow ratio (diff Qb/Qtot) for
a) identical parameter sets b) calibrated parameter sets used in the simulation runs

By shifting *ln(T*0*)* (lateral transmissivity when the soil is just saturated) according to the mean *ln(a/tanß)* values, one gets almost identical efficiencies and flow ratios. One main reason for the importance of *TO* in recalibration can be directly derived from basic equations used in TOPMODEL for the calculation of base flow. The results presented here indicate that discharge differences can largely be compensated by changing the mean transmissivity *T0*. After this compensation other discharge simulation results are almost independent of scale, supporting the results of Franchini et al., 1996; Wolock and McCabe 1995; Bruneau et al., 1995 and Quinn et al., 1995a).

3.3 Dynamics of Saturated Areas

Maps representing single pixel saturation counts show artificial chessboard-shaped patterns in large parts of the Stör and some parts of the Vils catchment (see Fig. 1).

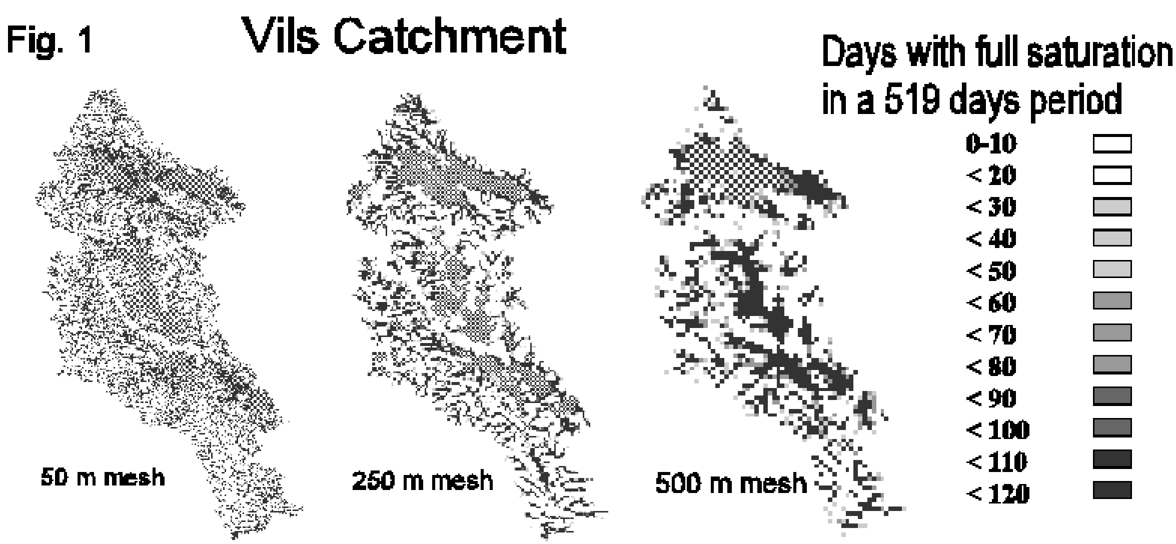


Fig. 1

All these areas characterized by flat orography are exposed to significant manipulation by the SINKS program. The patterns are due to assigning a gradient to flat area cells and are an indication for general problems in applying TOPMODEL, even though discharge curves are well reproduced.

Apart from this problem which is specific for flat regions, the simulation results can be summarized as follows:

Catchment averaged values for counts of full saturation are almost identical for all spatial resolutions. Differences can be explained by spatial aggregation and corresponding smoothing effects on the relief in the DEM and by changes in relief caused by the SINK program.

The highest saturation frequencies are independent of resolution as well. Saturation in converging zones increases if no CIT is specified.

In the case of the saturation patterns obtained for the Vils catchment with and without a CIT, linkages between some adjacent saturated pixels get lost when a CIT is specified. This fits the observation that increased CIT cause higher mean *ln(a/tanß)* values, which in turn enhance overland flow. Still, the maximum number of saturation counts stays the same if other parameter values are kept constant.

The saturation patterns obtained for the Buckener Au are comparable to soil patterns. Highest values for saturation counts occur where peaty soils are dominant, supporting observations of Merot et al. (1995).

The saturation maps of both the Frankenohe and the Vils catchment show that in steeper parts saturation is often limited to single grid cells. Due to missing spatially distributed data about topography-dependent soil moisture patterns for these catchments, the investigations here were simply focused on the question of, how moisture patterns based on high resolution DEMs differ from those obtained with low resolution DEMs. In case of large-scale TOPMODEL applications, heterogeneity observed for high resolution simulations cannot be represented in detail. However, for the Buckener Au at 110 m and 250 m resolution, the effects of small scale heterogeneity are better maintained since the subcatchment is orographically smoother.

4. Application of WET to the Elbe drainage basin

4.1 Model adaptation

While TOPMODEL was mainly used to investigate scaling issues, WET was applied to a large drainage basin as a whole. There are two main reasons for this strategy to use different modelling approaches for the different scales. Firstly, WET is more suitable for low temporal resolution (annual or monthly time steps). On the other hand, the correlation between temporal and spatial resolutions and scales is well known (Blöschl and Sivapalan, 1995). Further, it was possible to modify WET for use with spatially distributed data of both climate and soil, whereas the distribution version of TOPMODEL does not account for differences in any hydrological important feature except topography.

While TOPMODEL includes dynamics of different subsurface storages, an apriori assumption has been made for WET about vertical drainage from the unsaturated zone to deep ground water (through the deep drainage term D). This means, that WET is more intended to represent the overall influence of topography *on the longer-term hydrological behaviour of catchments* and different compartments of lateral flow. In addition, the scaling study with TOPMODEL demonstrated that patterns in soil moisture distribution can be maintained for coarser resolution through a simple modification of parameter values. By this, the application of WET to 1 km resolution data was justified. Still, modifications were needed to include the distributed climate and soil parameters.

First of all, the topographic index was calculated in GRASS (Fig. 2).

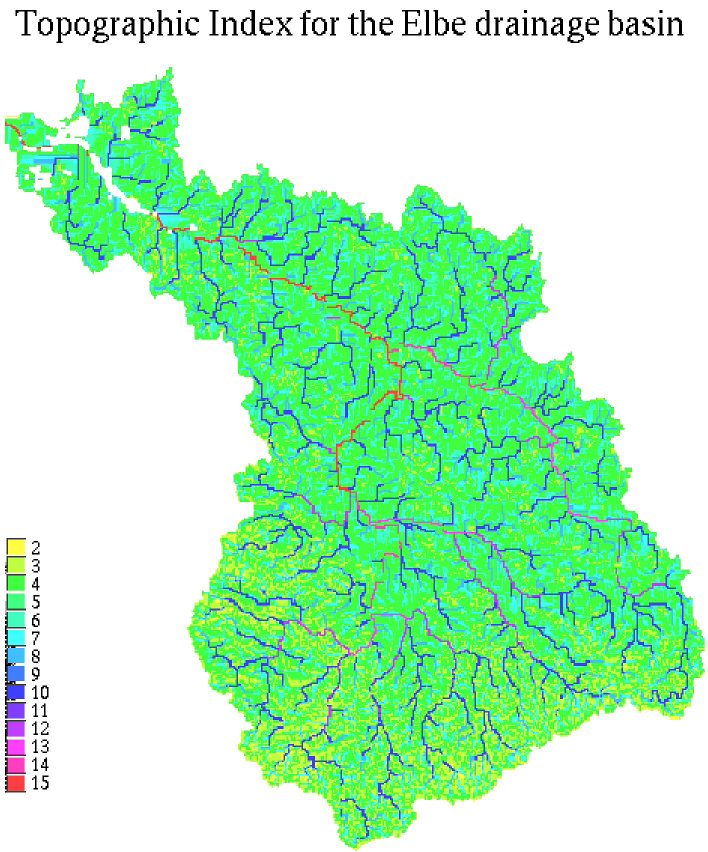


Fig. 2

The distribution function of the topographic index is comparable to those of other watershed studies (Quinn et al., 1995; Zheng et al., 1995) and the river network is well reproduced.

For such a large drainage basin it was necessary to account for spatially distributed precipitation and soil properties. On the basis of long-term monthly mean values of temperature and precipitation available for 48 meteorological stations in the region (years 1950-1980), spatially distributed monthly means of precipitation and temperature were calculated for every 1 x 1 km grid cell by the spline interpolation method (Hutchinson et al., 1993). Available soil water capacity was derived from soil maps. The program code was modified to include the distributed climate and soil parameters.

One crucial problem in the WET application was related to the actual evapotranspiration accounting for summer months, when potential evapotranspiration is quite high. The problem is directly related to the simplicity of the model, which does not take into account vegetation distribution, root depth, or root density. Including such additional information would improve the evapotranspiration component. On the other hand, the model should be kept as simple as possible for larger-scale applications. As a compromise, we used the soil moisture distribution of a previous month to get more reasonable results, restricting the actual evapotranspiration by precipitation plus change in soil moisture (following the water balance accounting procedure by Thornthwaite and Mather, 1957). By this, a quasi-dynamic element was introduced into the static equilibrium approach, which is reasonable for monthly time steps.

4.2 Results and comparison to previous studies

The results for the evapotranspiration and soil wetness index are quite reasonable. In April the mean monthly evapotranspiration is low and almost homogeneous for the whole region. In summer it is higher in the south, where precipitation and water-holding capacity of soils are higher. The distribution of the mean monthly soil wetness index differs essentially for winter and summer months, while the patterns for subsequent summer months are similar. A significant part of the territory is saturated (wetness index = 1) in winter, while certain areas are under water stress in summer (Fig. 3)

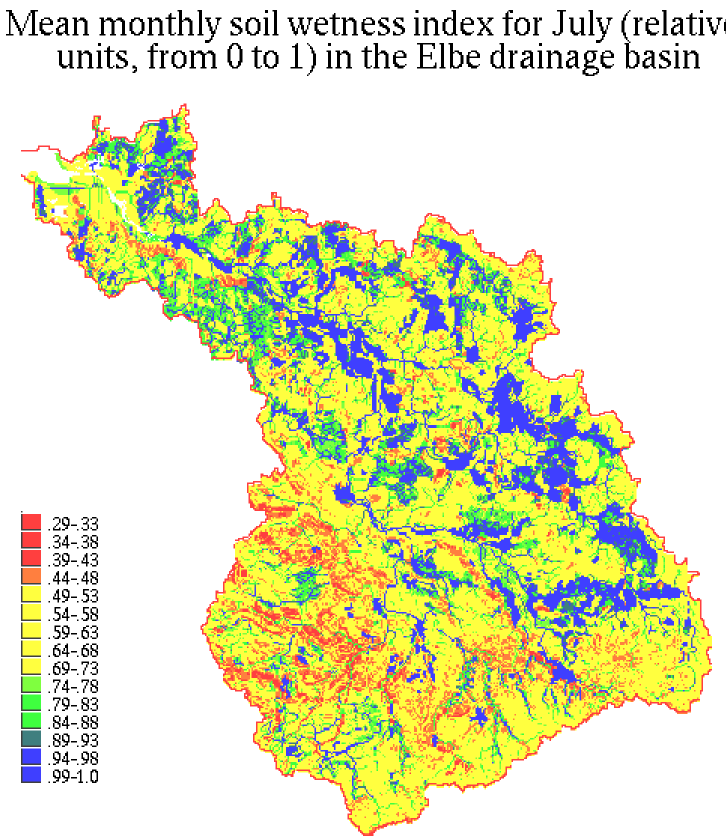


Fig. 3

Since WET does not calculate discharge time series, model validation must be performed on the basis of spatial patterns of subsurface moisture (i.e. soil moisture or ground water). However, neither field observations nor remote sensing methods provide data for the validation of results on soil moisture or evapotranspiration in the study region. Therefore, indirect methods had to be used. The results were compared to existing ground water maps (WASY GmbH, scale 1: 50000, see fig 4.) and some results of the independent terrestrial modelling performed in Germany (Wendland et al., 1993).

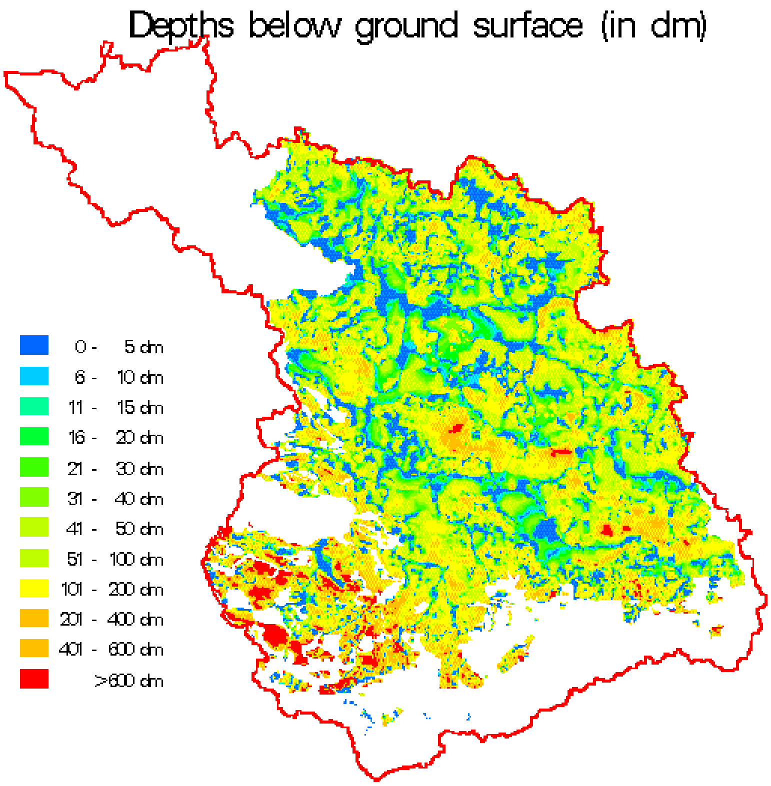


Fig. 4

The map of mean ground water depths for eastern Germany and the soil moisture index map generated by WET show similar patterns. Areas of high saturation occur mainly along riparian areas or zones of topographic convergence. On the other hand, dry areas defined by deep ground water tables and low wetness indices appear mainly in loess areas in the southern part of the basin. Comparison to the map of mean annual percolation rates (Wendland et al.,1993) was quite satisfactory as well. This indicates that a topography-based hydrologic approach focused on long-term dynamics is appropriate for time-averaged estimations of subsurface moisture patterns in a large region.

The distribution of the soil wetness index distribution in summer months (from May to August) was used to estimate soil moisture deficit in percent missing from the 50% of field capacity on average in summer months (Fig. 5).

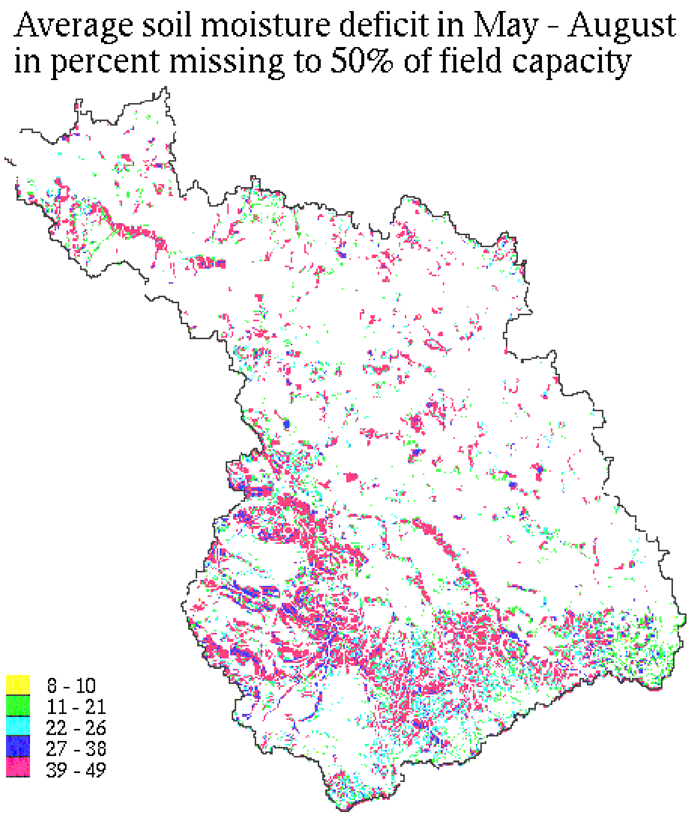


Fig. 5

Averaging the soil moisture deficit values for subbasins (areal average), we identified several classes of vulnerability, and, consequently, the vulnerable subbasins (classes 4 and 5, Fig. 6).

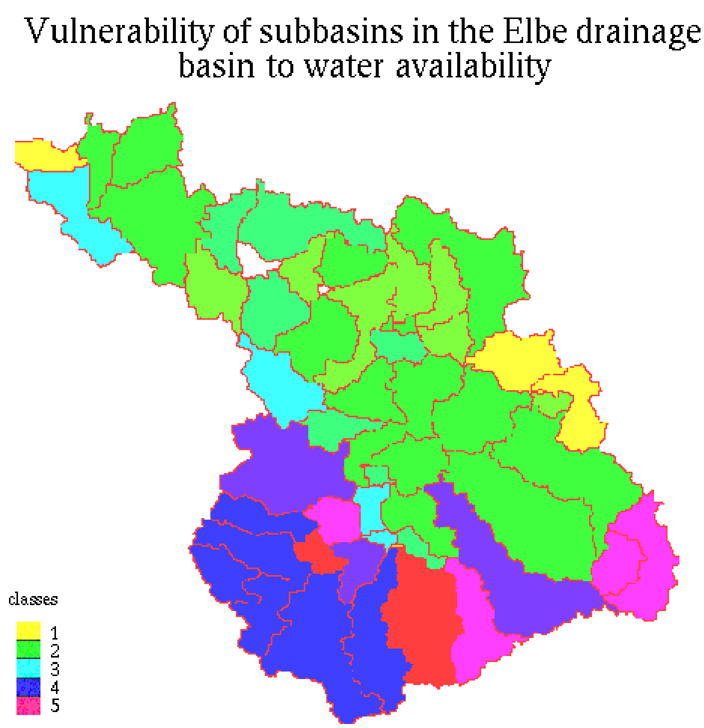


Fig. 6

5 . Conclusions

The TOPMODEL results show how differences in topography potentially cause differences in hydrologic patterns and soil moisture dynamics. They further show that the model may be applied to some lowland regions, if high resolution DEMs are available and the limits of application are carefully studied.

It was shown that runoff estimations are scale dependent and differences in discharge (mainly due to differences of mean *ln(a/tanß)* values) can to a large extent be compensated by changes in mean values of transmissivities *T*0, even for the 1000 m grid scale. After this compensation, spatially aggregated saturation patterns are more or less independent of scale for the chosen conditions.

Nevertheless, large scale DEM smoothing effects on topography may very much mislead the topography-dependent estimation of dynamic soil moisture patterns, especially if differences in vegetation and soil physical properties are not considered. It may be particularly important to account for intrapatch heterogeneity according to Avissar (1991) and Bruneau et al. (1995). In order to account for heterogeneity the high resolution *ln(a/tanß)* distribution functions should at least implicitly be included via assumptions based on the known relationship between these distributions for low and high resolution DEMs .

The approach described here allows critical subareas in relatively large drainage basins to be delineated. It can be used for preassessments in large-scale hydrological studies. Based on the knowledge on vulnerable areas, dynamic process-based models can then be applied to better understand the hydrological and biogeochemical processes and to reveal the feedback mechanisms of complex climate - biosphere interactions.

Our study demonstrates that even simplified models which include only the key factors determining system behaviour can be effective tools for the analysis and preassessment of the hydrological cycle in a watershed. A further improvement of equilibrium water balance modelling with monthly time steps would be possible for regions where distributed data on land cover and soil properties are available.

Acknowledgments

Thanks particularly to Paul Quinn (ADAS, Wolverhampton) and Keith Beven (Lancaster University, GB) who kindly provided the TOPMODEL, SINKS and GRIDATB codes and were always open for fruitful and constructive discussions. The TAPES-G package was obtained from the Centre for Resource and Environmental Studies, The Australian National University. We are grateful to Dr. J. Gallant for useful discussions.

The paper would not have been possible without the distributed climate data provided by Wolfgang Cramer (PIK), who also contributed through lots of useful comments.

The authors would like to thank Prof. Ripl (Free University of Berlin) and the Projektzentrum Ökosystemforschung, the "Landesamt für Wasser und Küste", the Univ. of the Armed Forces, Munich, and the German Weather Service (DWD) for providing data.

We are grateful to the Deutsche Forschungsgemeinschaft (DFG) and the Federal Ministry for Education, Science, Research and Technology (BMBF), which made the present work possible by providing financial support to the authors.

References

Arnell N. W. (1993). Data requirements for macroscale modelling of the hydrosphere. In: Wilkinson W.B. Macroscale modelling of the hydrosphere. *Proceedings of an international Symposium*, Yokohama, Japan 21 July 1993. IAHS Publications 214, 139-150.

Avissar, R. (1991). A statistical-dynamical aroach to parameterize subgrid-scale land surface heterogeneity in climate models. *Surv. Geophys*. 12, 155-178.

Becker, A., Krysanova V, Lahmer W. and Müller-Wohlfeil D.-I. (1993). Modelling of Hydrological and Hydrochemical Variablity under Environmental Change Impact. *Acta Geologica Hispanica*, No.2-3, 37-43.

Beven K. (1995). Linking parameters across scales: subgrid parameterizations and scale dependent hydrological models. *Hydrol. Process*., 9, 507-525.

Beven K.J. and Kirkby M.J. (1979). A physically based variable contributing area model of basin hydrology. *Hydrol. Sci.*

Bull., 24, 43-69.

Beven K.J. and Quinn P.F. (1994). Similarity and scale effects in the water balance of heterogeneous areas. In: *Proceedings of a AGMET Group Conference on The Balance of Water - present and Future* , Dublin, Ireland, pp. 69-86.

Blöschl G. and Sivapalan M. (1995). Scale Problems in Hydrological Modelling. *Hydrol. Process*., 9, 251-289.

Bruneau P., Gascuel-Odoux C., Robin P., Merot Ph., Beven K. (1995): Sensitivity to space and time resolution of a hydrological model using digital elevation data. *Hydrol. Process*., 9, 1, 69-81.

Chairat, S. and Delleur J.W. (1993). Effects of the topographic index distribution on predicted runoff using GRASS. *Water Resources Bulletin*, 29, 1029-1034.

Environmental Systems Research Institute, Inc. (ESRI) (1991). *ARC/INFO 6.0. User's guide*. Cell based modelling with GRID.

Famiglietti J.S. and Wood E.F. (1994). Multiscale modelling of spatially variable water and ebergy balance process. *Water Res. Research*, 30, 11, 3062-3078.

Franchini M., Wendling J., Obled C. and Todini E. (1996). Physical interpretation and sensitivity analysis of the TOPMODEL. *J. Hydrology* , 175, 293-338

Hutchinson M.F., Bischof R.J. (1993). A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales. *Austral. Met. Mag*. 31:179-184.

Leavesley G.H., Lichty R.W., Troutman B.M. and Saindon L.G. (1983). Precipitation-Runoff Modelling System: Users's manual. U.S. Geological Survey. *Water Resources Investigation Report*, 83/4238.

Merot Ph., Ezzahar B., Walter C. and Aurousseau P. (1995). Mapping waterlogging od soils using digital terrain models. *Hydrol. Process*. 9, 1, 27-34.

Moore I.D., Norton T.W. and Williams J.E. (1993). Modelling environmental heterogeneity in forested landscapes. J. Hydrol., 150, 717-747.Nash J.E. and Sutcliffe J.V. (1979). River flow forecasting through conceptional models, 1. A discussion of principles. *J. Hydrol*., 10, 282-290.

Nash J.E. and Sutcliffe J.V. (1979). River flow forecasting through conceptional models, 1. A discussion of principles. *J. Hydrol*., 10, 282-290.

O' Loughlin E.M (1981). Saturation regions in catchments and their relation to soil and topographic properties. *J. Hydrol*., 53, 229-246.

Priestley C.H.B. and Taylor R.J. (1972). On the assessment of surface space heat flux and evaporation using large scale porometers. *Mon. Weath. Rev*. 106, 81-92 .

Quinn P.F., Beven K.J. and Lamp R. (1995a). The In(a/tan ß) index: how to calculate it and how to use it within the TOPMODEL framework, *Hydrol. Process*. 9, 2, 161-182.

Quinn P.F., Beven K.J. and Culf A. (1995b). The introduction of macroscale hydrological complexity into land surface - atmosphere transfer models and the effect on planetary boundary layer development. *J. Hydrol*., 166, 421-445

Thornthwaite C.W. and Mather J.R. (1957). Instructions and tables for computing potential evapotranspiration and the water balance. *Publications in Climatology*, Laboratory of Climatology, vol 10, no 3

U.S.Army (1987). *GRASS reference manual*. USA CERL, Champaign, IL.

Wendland F., Albert H., Bach M., Schmidt R. (1993). *Atlas zum Nitratstrom in der Bundesrepublik Deutschland*, 96 S., Berlin: Springer-Verlag .

Wolock D. M. and McCabe, G. J. (1995). Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Res Research*, 31, 5, 1315-1324.

Wolock D.M. and Price C.V. (1994). Effects of digital elevation model and map scale and data resolution on a topography-based watershed model, *Water Res Research*, 30, 11, 3041-3052.

Zhang W. H. and Montgomery D.R. (1994). Digital elevation model grid size, landscape representation and hydrologic simulation. *Water Res. Research*, 30, 1019-1028.

Zheng D., Hunt, E.R. and Running, S.W. (1995). Comparison of available water capacity estimated from topography and soil series information. *Landscape Ecology*, in press.

D.-I. Müller-Wohlfeil, W. Lahmer, V. Krysanova & A. Becker
Potsdam Institute for Climate Impact Research (PIK)
P.O.Box 601203, Telegrafenberg, D-14412 Potsdam, Germany
Telephone: +49 331 288 2525 /-2524 /-2515 /-2541
Fax +49 331 288 2600
E-mail: dirkingm@pik-potsdam.de, lahmer@pik-potsdam.de, valen@pik-potsdam.de, becker@pik-potsdam.de

Gorokhovich, Yuri and Lorraine L. Janus

GIS Applications For Watershed Management.

Abstract.

The drinking water supply for New York City is protected by an aggressive watershed management program that intensified in 1991 with enactment of the federal Surface Water Treatment Rule. Watershed management strategies can most expediently be developed and guided by use of environmental models. Verified models are extremely important to the NYC water supply due to the large size and geographic variation of the watersheds, and the necessity to extrapolate, predict, and assess the impacts of change. Application of environmental models to develop watershed management strategies can be done through a GIS system. The main role of GIS is integration of different layers of information to produce model input. Several examples will illustrate GIS applications for the New York City (NYC) watersheds, including the Environmental Impact Statement (EIS) for Rules and Regulations, nutrient loading estimates and identification of turbidity sources.

Implementation of the NYC Rules and Regulations for water quality protection require EIS analyses. Through the GIS, different landuse and ownership data were merged with natural features such as streams, soils and lakes to produce a set of maps showing the potential areas that would be affected by implementation of the proposed watershed Rules and Regulations.

Nutrient control is an important aspect of water quality management. The Reckhow method of estimating the phosphorus loading is a lumped parameter model that was programmed in ARC/INFO. An interface was developed to link GIS datasets with equations that describe phosphorus loads. The interface was designed on the basis of "question-answer" communication which guides a user through a decision sequence that results in a basin specific loading estimate. It allows a user to display, calculate and modify parameters entered into the model to display the impacts of different scenarios of future land use.

Turbidity is another important parameter due to its interference with disinfection and as an indicator of other pollution. It is regulated by the Surface Water Treatment Rule (SWTR). A distributed parameter approach for modeling turbidity uses digital elevation model (DEM), hydro- geomorphological parameters of the stream channel, soil and hydraulic principles. Once developed, this model may guide best management practices (BMP) for erosion control.

Introduction

This paper describes GIS applications used for management of the watershed for the New York City Drinking Water Supply. Watershed management is of paramount importance to maintain high water quality and to avoid the need for costly purification systems. The extensive size of the watershed (approximately 1950 square miles) and varied geology, land cover, and land use make estimation and modeling procedures a necessity. The GIS is an invaluable tool in carrying out such procedures. The following is a description of how the NYCDEP uses its GIS for 1) the implementation of Rules and Regulations for the Protection from the Contamination, Degradation, and Pollution of the New York City Water Supply and its Sources, 2) the development of the phosphorus loading goals to prevent excessive eutrophication of the 19 reservoirs that comprise the water supply, and 3) the modeling of stream turbidity for source identification and management.

Rules and Regulations for Watershed Protection

New York City is in the process of promulgating new rules and regulations for watershed protection. The proposed watershed regulations are designed to minimize the discharge of pollutants into source waters from both point and nonpoint sources, to minimize the adverse impacts of erosion, to limit the discharge of nutrients into waters which may accelerate the eutrophication process of the reservoirs. It also provides for notification to the City's Department of Environmental Protection of ongoing or proposed activities which, either alone or in conjunction with other existing and proposed activities, may contaminate or degrade the City's water supply. Prior to promulgation of the Rules and Regulations, an Environmental Impact Statement (EIS) was conducted (during 1993 and 1994) to assess the environmental and socio-economic consequences of the proposed restrictions. The GIS was used extensively to conduct the analysis needed for the EIS.

The GIS allowed for the modeling and simulation of missing data, such as property boundaries, still providing necessary accuracy and functionality of the analysis, thereby overcoming the practical problems associated with incomplete data.

Landuse Data

Landuse areas are essential input for a number of environmental models, including the GWLF and Reckhow models. At the time that the EIS was conducted, there were no recent remote sensing landuse data available, so real estate property data (available as point files) were used instead. In New York State these data are available through the New York State Division of Equalization and Assessment (E&A). Since the E&A data provides only parcel centroids and not actual parcel boundaries, it was necessary to create simulations of parcel areas in order to obtain a spatial representation which would be useable for the various spatial analyses.

The simulation of parcel boundaries was accomplished through a method of interpolation called Thiessen or Voronoy polygons, (or proximal regions). This method is used when data have been collected at points, such as parcel centroids, and the analyst wishes to use area-based analytical techniques. The method consists of generating lines that join nearest neighbor points, then bisecting those lines with perpendicular mediators, and assembling the polygon edges using those lines. The whole procedure was programmed within ARC/INFO macro language code. Once the polygons were created, contiguous land uses were grouped together in clusters using the landuse code, extracted from the real estate attribute tables associated with the point files.

The GIS allowed for the modeling and simulation of missing data, such as property boundaries, still providing necessary accuracy and functionality of the analysis, thereby overcoming the practical problems associated with incomplete data.

Variable Buffer Model

A variable buffer is a spatial model that shows the limits around waterbodies and any other surface hydrological features within which no subsurface sewage treatment system may be built.The variables that change the shape of the buffer are soil types, soil hydraulic conductivity, slope or gradient of the terrain and fixed distances around streams and waterbodies. The GIS was programmed with an algorithm that created a variable buffer. This allowed comparison of experimental scenarios that included different combinations of soil and slope conditions.

Under the Proposed Watershed Regulations minimum buffer distances are set such that no part of any seepage unit or absorption field for a new subsurface sewage treatment system shall be located within 100 feet of a watercourse (stream) or wetland, or 300 feet of a reservoir or lake. By application of the variable buffer model, these limiting distances were found to vary from 100 to 1800 feet according to site specific soil and slope characteristics.

Developable Vacant Land Supply

In order to analyse the impacts of the Draft Watershed Regulations on development, the amount of existing vacant developable land in the watershed was determined using GIS. Projected growth for the period 1990-2010 was allocated into the developable vacant land, in order to evaluate the impacts the Draft Regulations would have on future growth. The developable land was defined as land that would be left open for development after the exclusion of all lands that have the following conditions: i) NYSDEC designated wetlands, ii) slopes over 25%, iii) public land, iv) shallow soils, v) land within the impervious surface limiting distance (100 ft. from watercourses and wetlands and 300 ft. from reservoirs, reservoir stems and controlled lakes, vi) parcels that fell completely within or had less than 1/4 acre outside the variable buffer for subsurface sewage discharge treatment systems, vii) clusters of less than one-half acre in places that are not sewered because of existing standards concerning minimum lot sizes for septic systems.

The GIS served as a framework to overlay the conditions mentioned above to produce maps showing developable land polygons. The analysis was performed to establish a baseline for the year 2010 and considering two ten years periods, 1990 to 2000 and 2000 to 2010, to determine potential displacement of new development due to the limiting distances under the Proposed Watershed Regulations. Maps were produced for the whole watershed area on mylars that could be used by planners in conjunction with USGS quads.

Nutrient Loading Estimates To Control Eutrophication

Nutrients control the biological production of the NYC reservoirs as demonstrated by the correlation between phosphorus and chlorophyll (NYCDEP, 1992). Excessive algal production leads to degradation of water quality and increased treatment. In order to control production, its necessary to estimate the current and critical phosphorus loads (Vollenweider, 1976). If the critical load is used as a goal, comparison with the current load indicates whether or not load reduction is needed.

A direct estimate of loading to a reservoir can be derived from information about its watershed characteristics. The Reckhow method to estimate the phosphorus loads is based on the concept that export coefficients are transferable and that two watersheds in the same region and with similar landuse patterns and geology will contribute the same loading of phosphorus per unit area. Estimates of the total annual mass of phosphorus entering a reservoir or lake is obtained by identifying landuse, applying export coefficients, then summing the annual phosphorus contribution for each nonpoint and point source within the watershed. By changing the assumptions on landuse and export coefficients, it is possible to evaluate the effects of future landuse changes on nutrient loadings, and subsequently water quality.

The GIS data which are used for the Reckhow model currently exist in a vector format which means that they are represented either by polygons, lines or points within ARCINFO vector format. ARCINFO is a GIS software that was developed by ESRI (Environmental Systems Research Institute). ARCINFO has a database module called INFO which provides the linkage between spatial objects and the database containing information about them.

There are 10 different data files that are linked together and provide necessary input parameters into the model interface:

Reservoir file: represents reservoir boundaries. Used for display only.

Precipitation file: represents meteorological stations as points. Attached attributes are mean yearly rainfall data (mm). Used to display station locations and estimate average precipitation for the selected basin.

Basin Boundary file: represents watershed basin boundaries. Used for the landuse analysis. It was overlayed by the landuse data to produce a table showing areas of different landuse categories within each basin.

Landuse file: represents polygons attributed by landuse categories. For the Reckhow model only 4 categories were used: agricultural, urban, water and residential.

House file: represents houses as points. Used in the model to select houses within certain distance from the streams.

Sewer data file: represents polygons that are sewer district boundaries. This file was used to exclude houses that are located inside the sewer districts.

Soil file: represents polygons with soil type attributes. Ideally each soil type should have had phosphorus retention coefficient value. In a current version user can assign only one value per all soil types. This coefficient then applied to the amount of phosphorus released from houses.

Sewage Treatment Plants file: represents points that have such attributes as a plant name and yearly phosphorus loading values for both hydrological and a regular year.

Reservoir info file: this is a file formatted within INFO database. It has data on residence time (year) and depth (m) for both hydrologic (H) and regular (R) years for all reservoirs and used also to select the basin.

Extra info file: this is a file formatted within INFO database and it has values representing additional phosphorus inputs into specific reservoirs.

An interface was developed to link the GIS datasets with the Reckhow model equations that describe phosphorus loads. The Reckhow phosphorus loading, lumped parameter model was fully programmed within ARC/INFO. This interface allows a user to display, calculate and modify parameters entered into the model to display different scenarios of current and future land use. The interface was designed on the basis of "question-answer" communication which guides a user through all necessary steps. The following is a brief summary of the steps:

1. Selection of the year to model (hydrological or regular); 2. Selection of the watershed basin to model. 3. Selection of the precipitation (enter your own value or select from the available precipitation stations shown on the screen). 4. Selection of the export coefficients. 5. Selection of the parameters for the septic calculations. 6. Spatial analysis of the sewage treatment plant (STP) data. 7. Select a phosphorus release value (actual or which would be under restrictions). 8. Option to edit release values. 9. Select soil retention coefficients. 10. Spatial analysis of the potential septic data (amount of houses outside chosen buffer area around streams, wetlands, lakes and reservoirs). 11. If applicapable, program automatically includes additional phosphorus inputs into the reservoir from the data file. 12. Final report output.

The final report consists of a summary listing of phosphorus sources and their contributions to the total load displayed in the following format:

Phosphorus Source........ Phosphorus Load (kg/year)

Agricultural Land 2215

Forest Land 1032

Urban or Built-up Land 285

Water 387

OTHER SOURCES:

Household Septics.........................6251

Sewage Treatment Plants 144

Other Point Sources 9857

TOTAL...................................20171

This table allows one to identify a relative importance of different sources and can guide the selection of remediation measures.

Sub-basin Delineation for Export Coefficient Development:

Reckhow model utilizes export coefficients that are calculated from hydrological flow and stream nutrient concentrations. Flow and concentration measurements are converted to total mass loadings. Mass loadings are then divided by sub-basin areas to calculate export coefficients, therefore, it is necessary to know the size of the area draining to the site. The GIS system allows automatic processing of the area delineation. To accomplish this, elevation data were converted into the Digital Elevation Model (DEM). Then DEM file was converted into the raster grid containing information about the flow direction of the water along the surface of the land. When the location of the sampling site was entered into the GIS system, software automatically delineated the boundary of the area draining into that location.

Turbidity Source Study to Minimize Disinfection Treatment, Erosion, and Reservoir Capacity Loss

Turbidity is another important parameter influencing water quality. According to EPA regulations (40 CFR Parts 141 and 142, 1989) the turbidity level cannot exceed 5 NTU in representative samples of water immediately prior to the first or only point of disinfection. These regulations make turbidity one of the most important factors in water quality monitoring, because it reduces disinfection performance and increases treatment costs. In addition soil loss and the resulting sedimentation in the reservoirs can reduce their storage capacity.

New York City currently carries out an extensive field sampling program to identify sources of turbidity. Nonetheless, logistics and manpower limit investigations to only a few of the many sub-basins within the watershed. The goal of the current investigations is to pinpoint places on streams that would most benefit from the implementation of best management practices. Although field sampling allows us to address a few specific locations, modeling is needed to identify turbidity sources throughout the extensive watershed area.

Modeling of turbidity is a future goal. Most of the current modeling approaches are based on the available lumped parameter Universal Soil Loss Equation (USLE). The approach here is to create a distributed parameter model. This approach will utilize soil potential for delivering particles and sediment transport equations describing soil particle movement along stream channels. Turbidity is an excellent parameter for modeling because it is easily measured. The future GIS application will include: i) use of the digital elevation model to derive potential energy levels in the water proportional to the topographic slope, ii) coding stream segments by their predicted kinetic energy levels, (iii) use of soils and surface geology classify stream segments according to their potential to deliver soil particles and, iv) estimation of the suspended material and turbidity from these parameters that can then be compared to field measurements. The practical application of this work will be identification of reaches of stream that should be examined as candidates for best management practices to reduce turbidity.

There are two main steps in producing final turbidity maps based on the proposed model. These two steps are the dynamic segmentation of streams and subsequent graphical representation of the results. Dynamic segmentation allows calculation of necessary parameters (such as velocities, particle concentrations, discharges, etc.) along the stream network. These parameters serve as input for the model that then calculates different turbidity levels for each segment. Graphical representation allows model results to be displayed as hardcopy maps, monitor images, animation series, etc.

Dynamic segmentation is a method for creating, editing, querying and displaying attributes of linear features without changing their actual topology. In the context of hydrology, linear features are streams and their network. Practical implementation of the dynamic segmentation for the turbidity modeling requires the following: i) definition of sequential routes within the stream network where each route is assigned a unique number, and ii) define the route-measure system i.e. a standard table that has "from" and "to" distances. These distances define segments along stream channels and each segment has hydro- geomorphological attributes that describe the channel morphology and potential stream power energy. The measure-route table is then linked with the stream routes as a one-to-many relational database.

After preparing dynamic segmentation framework, any value can be extracted from the table and used for further calculations. Graphical representation of different turbidity levels was done using the `buffer' algorithm. According to the observed range of turbidity values six main categories were defined: < 10, 10 - 50, 50 - 100, 100 - 150, 150 - 200 and > 200. These categories were numbered 1 to 6. Turbidity values, stored in the data file were converted into six buffer categories. The `buffer' module can then be used to produce a digital dataset of the stream buffer width that correspond to different turbidity levels. The digital dataset can then be displayed. An application of this mapping could be animation of a storm event by producing a series of maps showing turbidity changes that occur during the progress of a storm.

Conclusions

The GIS is invaluable tool for watershed management as illustrated by the examples above. The applications are diverse, ranging from data development to environmental modeling. Data development for the watershed management includes such typical GIS applications as spatial overlay of existing GIS data to develop new layers and transformation of topological entities (i.e., points or lines) to polygons. Environmental modeling for watershed management is evolving towards distributed parameter models that can be supported by such GIS tools as dynamic segmentation and mapping. Distributed parameter models allow for better geographic resolution of pollution sources, and this will lead to efficient application of best management practices and other watershed protection measures.

Acknowledgments

There were many people within DWQC who devoted their time and efforts to the collection and organization of necessary information to produce the GIS and other data for the modeling and mapping projects. In particular, we appreciate the cooperation of Ms. Esther Siskind, Project Manager for the Proposed Rules and Regulations; Ms. Cheryl Weisner who produced all final maps for the Proposed Rules and Regulations and programmed many ARC/INFO macro modules; and Dr. Kim Kane, whose enthusiasm and organizational efforts expedited the nutrient modeling task.

References:

NYCDEP, 1992. New York City Drinking Water Quality Control 1992 Annual Report. NYC-DEP, DWQC and Sources Report. October 1993.

NYCDEP. 1993. Final Generic Environmental Impact Statement for the Proposed Watershed Regulations for the Protection from Contamination, Degradation, and Pollution of the New York City Water Supply and Its Sources. Report, NYCDEP, vol. I.

NYCDEP. 1993. Final Generic Environmental Impact Statement for the Proposed Watershed Regulations for the Protection from Contamination, Degradation, and Pollution of the New York City Water Supply and Its Sources. Report, NYCDEP, vol.II.

Reckhow, K.H., M.N.Beaulac, and J.T.Simpson. 1980. Modeling phosphorus loading and lake response under uncertainty: a manual and compilation of export coefficients. U.S.Environmental Protection Agency (EPA 440/5-80-011). 214 pp.

Vollenweider. 1976. Advances in Defining Critical Loading Levels for Phosphorus in Lake Eutrophication. Mem. Ist. Ital. Idrobiol. 33: 53-83.

Author Information (name, title, organiz., tel., FAX, e-mail).

Yuri Gorokhovich, Graduate Center, CUNY

Earth & Environmental Sciences Program

33 West 42 Street, NY 10036-8099

Tel: (914)742-2085; FAX: (914)773-0365

E-mail: yuri@valgis.dep.nyc.ny.us

Dr. Lorraine L. Janus, Deputy Chief, DWQC

New York City Department of Environmental Protection, 465 Columbus Ave., Valhalla, NY 10595

Tel: (914)742-2082; FAX: (914)773-0365

Randy Hills[1], Ann McManamon[1], and Robert K. Hartman[1]

Snow Estimation and Updating System (SEUS)

ABSTRACT

The National Weather Service Office of Hydrology has developed a methodology to generate real-time, gridded snow water equivalent estimates using ground-based and airborne snow data collected over the Western United States. The gridded snow water equivalent estimates incorporate the spatial variability of the snowpack induced by the orographic effect in the West. The Snow Estimation and Updating System (SEUS) uses a geographic information system to store, analyze and display the spatial data necessary to perform the estimation. The gridded information is used to derive snowmelt characteristics and to develop long-term mean snow water equivalent data. A conceptual hydrologic model is used in the development of many of the parameters needed during this calibration step. The point snow water equivalent data are interpolated into a gridded product using data derived during the calibration step. Basin boundaries are used to identify the area included within each basin so that the gridded data can be analyzed to determine the average snow water equivalent over subareas within each basin. The estimates are used to update the snow water equivalent states of National Weather Service River Forecast System snow model.

The system is presently implemented in the Colorado River Basin, portions of the Sierra Nevada, and portions of the Columbia River Basin.

INTRODUCTION

In the Western United States, a significant portion of the annual runoff is generated from snowmelt. Accurate estimates of the snowpack can greatly enhance seasonal water supply forecasts, which are useful in estimating hydropower generation, planning reservoir releases, and determining water allocations. Regression techniques work well in estimating snowmelt in average years, however, these techniques tend to be inaccurate in extreme years.

Water supply is also forecasted with conceptual models. The National Weather Service uses the Extended Streamflow Prediction (ESP) technique as part of the NWS River Forecast System (NWSRFS) to perform long-term forecasts of streamflow. ESP uses current streamflow, soil moisture, and snowpack conditions along with historical time series of precipitation and temperature to estimate future streamflow conditions. Based on the likelihood of future precipitation and temperature time series, the resulting streamflow hydrographs can be analyzed to produce probabilistic forecasts of streamflow peaks, volumes, etc.

Because of the difficulty in accurately estimating precipitation in the mountains, the estimates of the initial conditions provided to the models are often inaccurate. The reliability of ESP forecasts can be increased by using snow water equivalent observations to update model-simulated snow cover conditions. The Snow Estimation and Updating System (SEUS), developed by the National Weather Service, was created for this task. The SEUS utilizes existing ground and airborne snow water equivalent observations, providing better estimates of snowpack conditions for making water supply forecasts (Day, 1990).

SNOW ESTIMATION AND UPDATING SYSTEM (SEUS)

SEUS consists of four components: calibration, operational, updating, and administration. The calibration component analyzes historical snow observation data and develops the parameters needed to estimate snow water equivalent operationally. The operational component utilizes these calibration parameters, along with near real-time snow observation data, to determine gridded snow water equivalent. The updating component computes new snow water equivalent states for the conceptual snow model based on the weighted contributions of the historic simulated snow states and the estimates of the snow states developed using historic snow observations. The administration component manages individual user calibration and operational data.

Gridded, line, and point data created and used by SEUS is managed by the Geographic Resources Analysis Support System (GRASS) GIS. GRASS is a raster-based public-domain GIS developed for UNIX platforms by the U.S. Army Construction Engineering Research Laboratory (USACERL). GRASS was chosen in the development of SEUS because new functions can be added easily with scripts utilizing GRASS commands or with new commands utilizing GRASS C library routines. SEUS calibration, operational, and administrative processes are simplified by providing a graphical user interface for the user.

For purposes of illustration, the basin "Kings River at Pine Flats", located on the western slope of the Sierra Nevadas in central California (Figure 1), will be used to demonstrate different outputs of SEUS calibration, operation, and updating components. This basin has a drainage area of approximately 1500 square miles, ranges from 900 to 13100 feet in elevation, and averages approximately 32 inches of winter precipitation.

Figure 1. Basin "Kings River at Pine Flats".

CALIBRATION COMPONENT

The purpose of SEUS is to interpolate point snow water equivalent data into gridded estimates of snow water equivalent. A direct interpolation of point snow water equivalent in the West, however, does not account for orographic effects. One possible method to account for these effects is to transform point snow water equivalents into point standardized deviates using the following equation:

$$Z = \frac{(x - \overline{x})}{\sigma}$$

where,

Z = standardized deviate,
x = snow water equivalent observation,
$\overline{x}$ = long-term temporal mean,
σ = long-term temporal standard deviation,

A gridded snow water equivalent estimate is created by correlating points based on their distance from each other. This requires an estimate of the spatial correlation function of the standardized data. A correlation function was developed for each basin for the first of each month of the snow season using historical station data. This function is expressed as the correlation between each station pair as a function of distance. The equation

$$\rho = ce^{-dx}$$

where

ρ	=	correlation coefficient,
c,d	=	regression coefficients,
x	=	distance between points

is used to fit the data. Figure 2 illustrates this relationship for the Kings River at Pine Flats for the month of January.

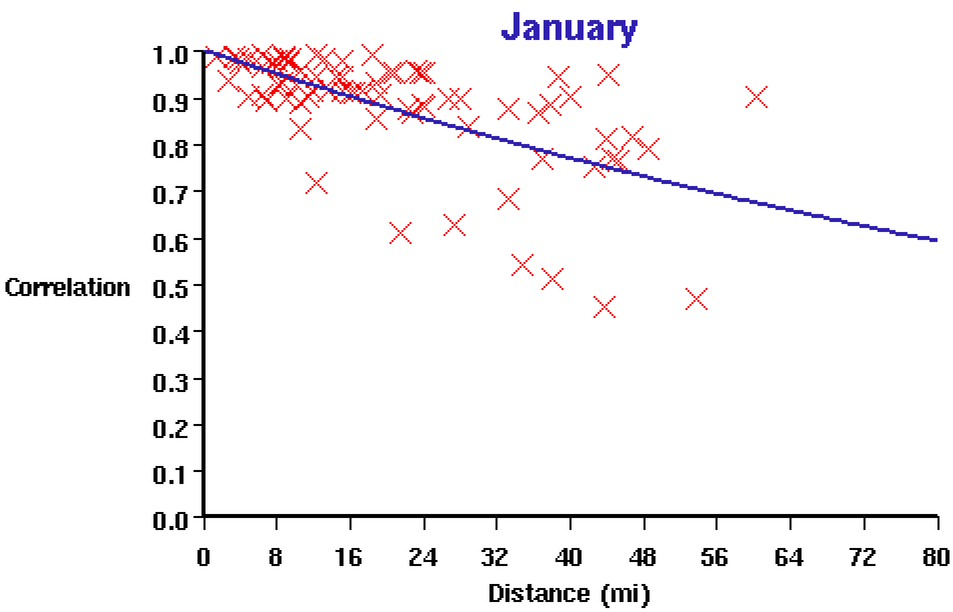


Figure 2. Correlation-distance function for Kings River at Pine Flats.

Given station means and standard deviations and a basin correlation function, point snow water equivalents can be transformed into standardized deviates and a gridded interpolated field of standardized deviates can be produced. However, a gridded field of snow water equivalent is really desired. To transform the standardized deviate grid into a snow water equivalent grid, an estimate of the mean and standard deviation of the snow water equivalent at each grid point is needed.

Estimates of the standard deviation of snow water equivalent at a grid point are formed from a mean-standard deviation relationship derived from historical snow water equivalent data for the first of each month of the snow season. The form of this relationship is assumed to be:

$$\sigma = a\bar{x}^{b}$$

where

σ	=	standard deviation,
a,b	=	regression coefficients, and
$\bar{x}$	=	mean.

Figure 3 illustrates this relationship for the Kings River at Pine Flats for the month of January.

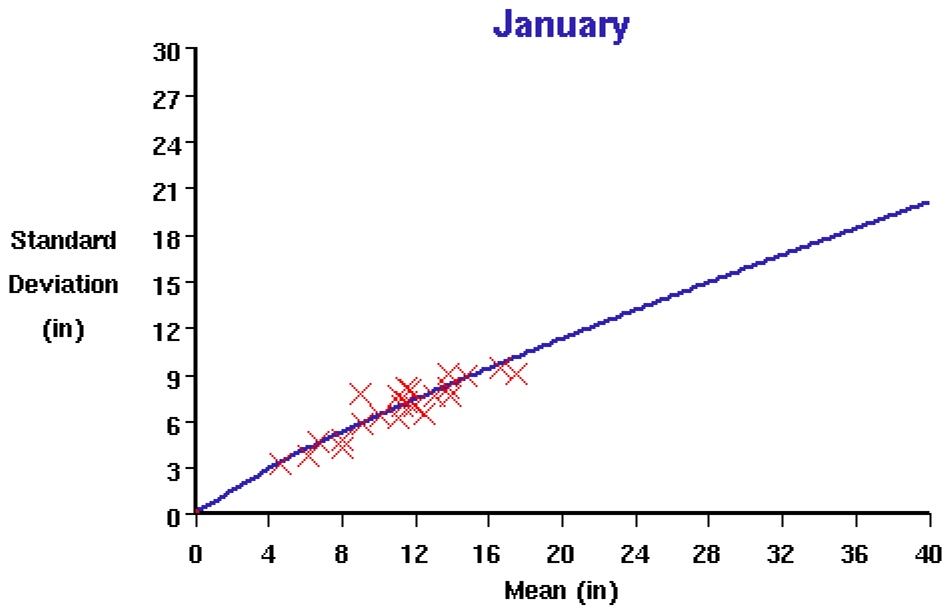


Figure 3. Mean-standard deviation relationship for Kings River at Pine Flats.

Using this relationship, transforming a standardized deviate grid into snow water equivalent grid is now solely a function of the mean snow water equivalent at a point.

MEAN MAP COMPUTATION

Estimates of the mean snow water equivalent at a grid point are derived by modeling snow accumulation and ablation, taking into account the precipitation and site characteristics of the grid point. The GIS and an existing, calibrated, NWSRFS snow model are used to estimate gridded snow water equivalent weekly through the snow season.

Modeling mean snow water equivalent at individual grid points could be extremely computationally intensive given the size of a basin, the grid resolution, and the length of the historical record. As a way of expediently modeling mean snow water equivalent, grid points are lumped into zones based on common snow melt characteristics. First, melt factor classes are formed as a function of aspect, slope, and forest cover. Aspect and slope, both computed from digital elevation data, are combined to form a new surface representing an index of available solar radiation. East and west-facing slopes are assumed to receive the same amount of solar radiation over a day as a horizontal surface. Consequently, for the purposes of SEUS, the available solar radiation is represented by three classes: north, south, and horizontal. Vegetation data is classified into forested and open areas. These solar radiation and vegetation classes produce six melt factor classes.

Melt at a grid point is a function of temperature as well as melt factor. Since temperature is well correlated with elevation, temperature-induced melt is determined by choosing representative elevations within the basin. The GIS is used to derive the range of elevations within the basin, and along with subarea cutoff elevations from the NWSRFS basin model, representative elevations are chosen which define the range of elevation data.

Given these melt factor classes, the list of representative elevations, and the mean areal temperature (MAT) and mean areal precipitation (MAP) time series for the basin, the NWSRFS snow model can compute a mean snow water equivalent for a point within a basin. The melt factors in the snow model are adjusted to represent each melt factor class, the MAT time series is lapsed from its representative elevation to match each required elevation within the basin, and the MAP time series is adjusted to match expected precipitation amounts within the basin. The resulting simulations are used to define a relationship for the melt factor class, mean seasonal precipitation, and mean snow water equivalent weekly from January through June. Mean snow water equivalent surfaces are then produced from these relationships utilizing the snow melt factor surface, the long-term mean October through April precipitation surface, and the elevation surface as shown in Figure 4 for Kings River at Pine Flats for January 22.

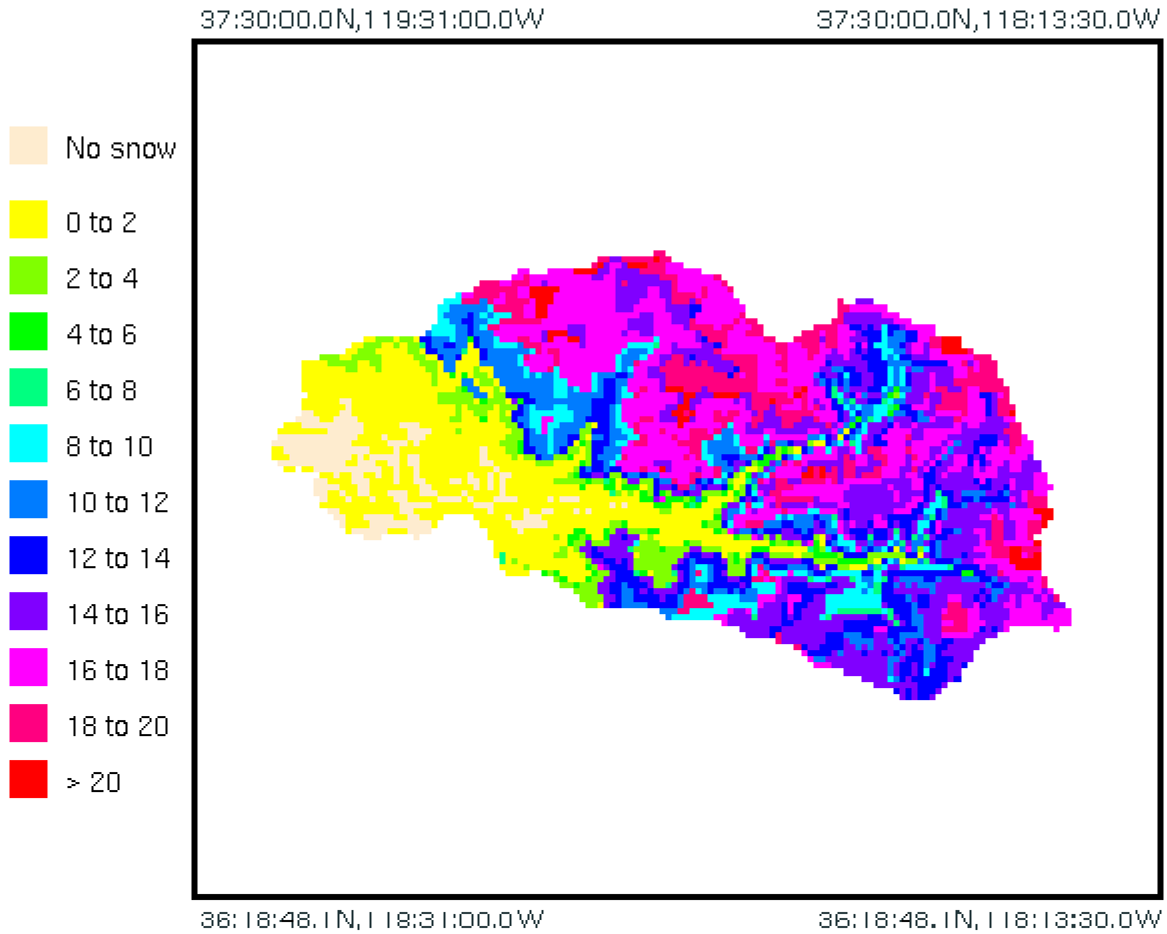


Figure 4. Mean snow water equivalent surface (in inches) for Kings River at Pine Flats for January 22.

OPERATIONAL COMPONENT

A gridded field of snow water equivalent is computed by first creating a gridded field of standardized deviates. This field is developed using the standardized deviate definition, snow water equivalent observations, the basin's correlation-distance function, and an interpolation procedure detailed in Day (1990). The interpolation routine uses all available observations in a pre-defined area within and around each basin (previously, a maximum of twenty observations were used), computing the standardized deviate for each grid point within the basin. A standardized deviate field for Kings River at Pine Flats for January 20, 1995, is shown in Figure 5.

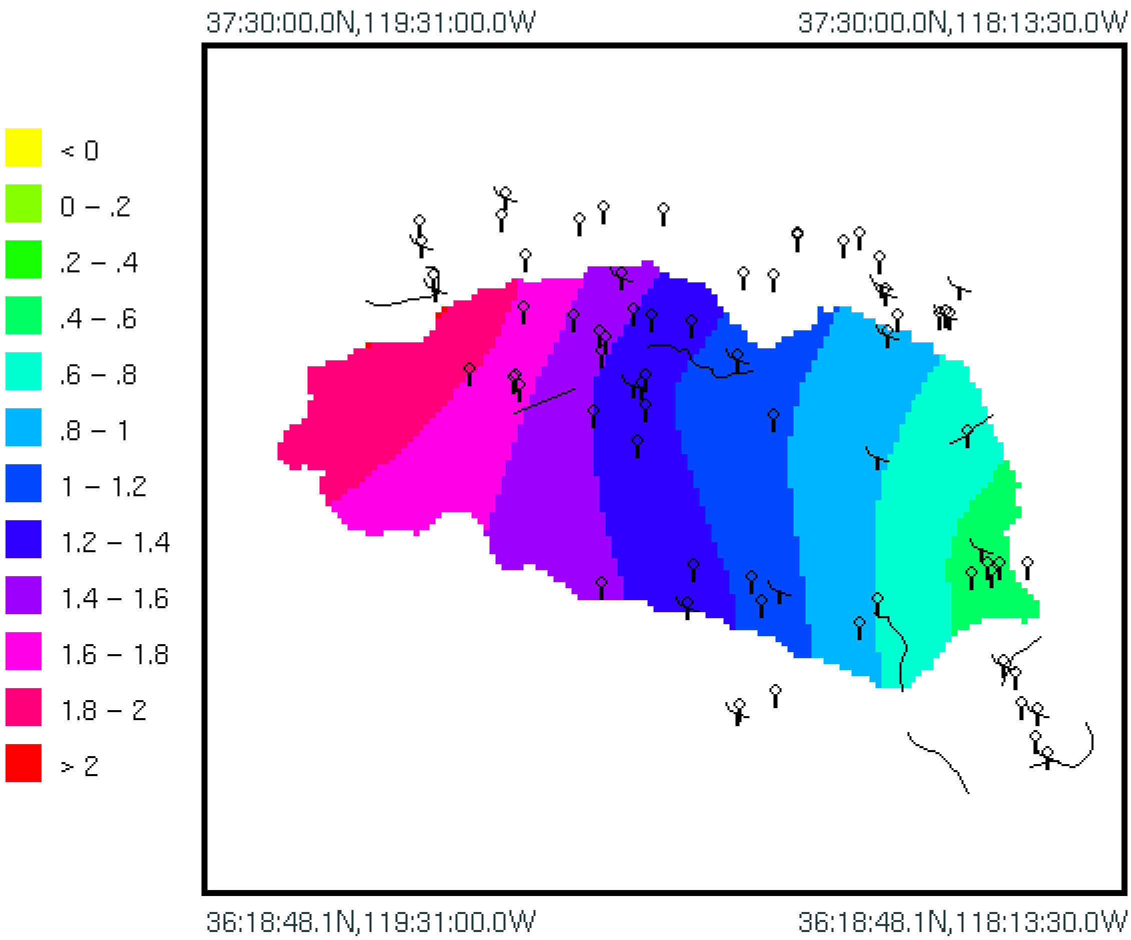


Figure 5. Standardized deviate surface computed for Kings River at Pine Flats for January 20, 1995.

From this gridded field of standardized deviates, a gridded field of snow water equivalent is created by using the basin's relationship between point long-term means and standard deviations and the equation defining standardized deviates, recasting it as

$$x = Z\sigma - \overline{x}$$

or

$$x = Z * a\overline{x}^b - \overline{x}$$

to determine snow water equivalent. A snow water equivalent field for Kings River at Pine

Flats for January 20, 1995, is shown in Figure 6.

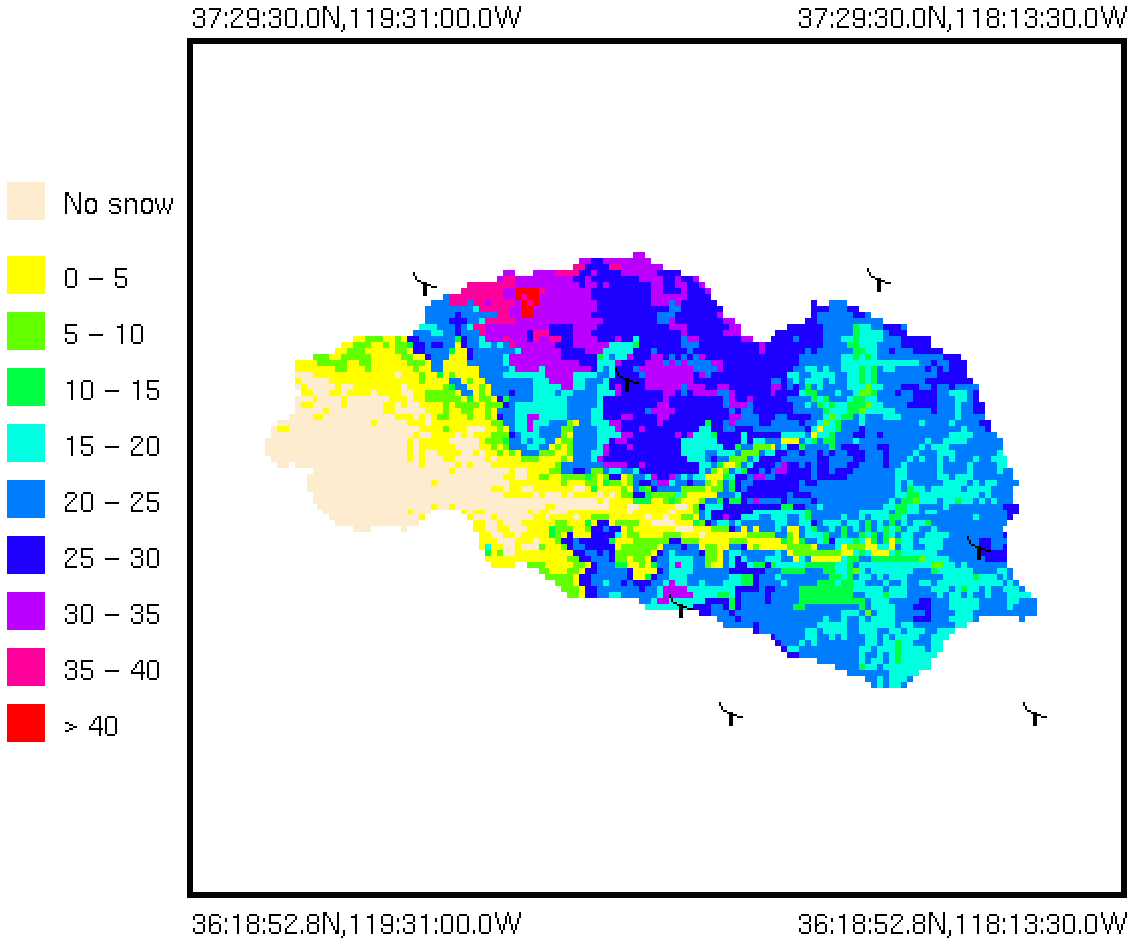


Figure 6. Snow water equivalent surface (in inches) computed for Kings River at Pine Flats for January 20, 1995.

The operational system can be run either in batch or interactive mode at any time between January 1 and June 22. Typically, the system is first run in batch mode for all the basins, and the resulting standardized deviate and snow water equivalent surfaces are examined for abnormalities. The system is then run interactively for these questionable basins, and the user can examine the snow water equivalent observations and discard those which may be inappropriately biasing the interpolation. These basins can be rerun, and once the user is satisfied with the new results, the user can combine these with the existing maps.

UPDATING COMPONENT

All of the information needed to estimate snow water equivalent is now available, however, the snow water equivalent estimated using the interpolation procedure may not be consistent with the snow water equivalent states in the conceptual snow model. Historical estimates of the snow water equivalent needed by the model are generated by computing the model states which would have been necessary on a specific date in order for the model to simulate the seasonal runoff that was actually observed. These estimates are called pseudo-observed snow water equivalent, and they represent the best estimate of the optimal snow water equivalent model states. Pseudo-observed snow water equivalents are developed for each basin subarea which has been calibrated for NWSRFS.

In order to account for biases between the pseudo-observed and the estimates of snow water equivalent form the interpolation procedure, regression relationships are developed from the historical data. Pseudo-observed values are estimated for the first of each month for the entire historical record. Similarly, the interpolation procedure is performed for the first of each month throughout the historical record. The GIS is used to compute basin subarea averages from the gridded estimates of snow water equivalent. Regression relationships, which predict pseudo-observed values from basin average snow water equivalent, are developed for the first of each month (Figure 7 illustrates the relationship for Kings River at Pine Flats for first of February). These relationships are used in the operational system to compute estimates of the model snow water equivalent states that can be used for updating.

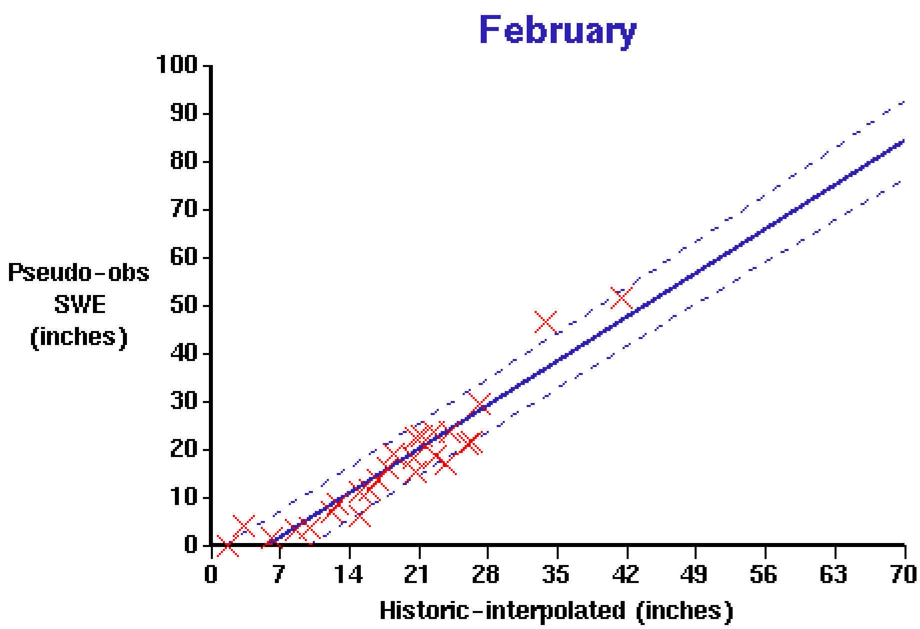


Figure 7. Pseudo-observed versus historic-interpolated snow water equivalent relationship for Kings River at Pine Flats.

FUTURE WORK

The NWS is adding a new snow operation, SNOW-43, to NWSRFS to enhance the updating capability of the current snow accumulation and ablation model, SNOW-17. SNOW-43, a state-space version of SNOW-17, uses a Kalman filtering updating procedure which optimally combines observed snow water equivalent estimates from SEUS with simulated states generated by the SNOW-17 model. The procedure uses error estimates from both processes to update the snow model states. SEUS will be modified to compute an error estimate of a subarea mean snow water equivalent value along with the currently estimated subarea mean snow water equivalent.

The correlation between snow water equivalent observations of any two sites can be a function of other factors besides distance. Other parameters, such as elevation or available solar flux, could also influence the correlation between two stations. A multiple linear regression technique is being examined to account for these additional factors.

An increase in the number of melt factor classes is being investigated. Adding more classes, besides the currently used north, south, and horizontal classes, may provide a better definition of melt. Using forest density, in conjunction with forest cover type, will also be investigated.

A verification system will be implemented to quantify the improvement of runoff forecasts due to the SEUS updating procedure. The method will incorporate techniques developed as part of the National Weather Service Extended Streamflow Prediction (ESP) procedure.

REFERENCES

Day, Gerald N., "A Methodology for Updating a Conceptual Snow Model with Snow Measurements", NOAA Technical Report NWS 43, Department of Commerce, Silver Spring, MD March 1990.

Day, Gerald N., L. E. Brazil, C. S. McCarthy, and D. P. Laurine, "Verification of National Weather Service Expanded Streamflow Prediction Procedure", Proc. on Managing Water Resources During Global Change, American Water Resources Association Symposium, Reno, NV, ASCE, 1992.

Geographic Resources Analysis Support System (GRASS) User's Reference Manual, Version 4.1, U.S. Army Corps of Engineers Construction Engineering Research Laboratory, Champaign, IL, 1993.

McManamon, A. , Szeliga, T. L., Hartman, R. K., Day, G. N., and Carroll, T. R., "Gridded Snow Water Equivalent Estimation Using Ground-based and Airborne Snow Data", Proceedings of Eastern Snow Conference, Quebec City, Quebec, pp 75-81, 1993.

---

[1]*National Operational Hydrologic Remote Sensing Center*
Office of Hydrology, National Weather Service
1735 Lake Drive West
Chanhassen, Minnesota 55317-8582
Telephone (612)361-6610, Facsimile (612)361-6634
email: {rhills,amcmanamon,rhartman}@nohrsc.nws.gov
www: http://www.nohrsc.nws.gov

Robert K. Hartman[1], Andrew A. Rost[1], Donald M. Anderson[1]

Spatial Distribution of Snow Water Equivalent Observations in Mountainous Terrain

ABSTRACT

Quantitative assessments of the spatial distribution of snow water equivalent (SWE) provide valuable insight and information for water management, hydrologic forecasting, and emergency preparedness entities throughout the United States and Canada.

The National Weather Service and its cooperators, principally the Natural Resources Conservation Service and California Department of Water Resources, gather SWE and related data for approximately 2500 fixed sites in the western United States and a portion of southwest Canada. Observation sites are exclusively located in mountainous regions and predominately at moderate elevation. As such, the lower and higher elevation ranges are not well represented. Satellite imagery can be used to augment SWE field observation sites, particularly at lower elevations. Analysis of classified imagery (i.e., snow, ground, and cloud) permits an estimate of the spatial distribution of the snow line (zero SWE interface). Gridded estimates of SWE can be obtained by combining the satellite-derived classification and snow line with field observations of SWE in a process that incorporates elevational detrending.

INTRODUCTION

Near real-time quantitative assessments of areal extent of snow cover and gridded snow water equivalent (SWE) provide valuable insight and information for water management, hydrologic forecasting, and emergency preparedness agencies throughout the United States and Canada. Since 1986, the National Operational Hydrologic Remote Sensing Center (NOHRSC) of the National Weather Service (NWS) Office of Hydrology (OH) has provided these data in rasters and alphanumeric tabulations generated using satellite image processing, spatial interpolation, and GIS analysis. Until recently, each of these processes was performed independently. They are now largely integrated in the newly developed NOHRSC Operational Product Processing System (OPPS) (Hartman et al., 1995). This paper describes the process by which gridded snow water equivalent estimates are generated in mountainous terrain.

BACKGROUND

Snow water equivalent at a point is a function of many physiographic factors. Dominant factors include elevation, orientation (slope and aspect), vegetative cover, exposure, and weather. Snow accumulation and snow melt in rugged terrain can be viewed as a surface exhibiting high frequency characteristics. Influences interact at a continuum of scales-from very small to quite large-to generate the snow cover we see in the mountains each winter.

In an attempt to quantify the dominant factors that influence the snowpack in the Western U.S., the National Weather Service developed the Snow Estimation and Updating System (SEUS). The purpose of SEUS is to estimate a gridded SWE field that can be used to update a process simulation runoff model resulting in improved streamflow forecasts (Day, 1990, McManamon et.al, 1993). While the accuracy of individual pixel estimates from SEUS may be questionable, the values are reasonable when integrated within a hydrologic basin and provide for improved streamflow simulation. The process of SEUS implementation is somewhat laborious, and to date, only about twenty-five percent of the mountainous Western U.S. can be estimated.

In the meantime, gridded estimates of snow water equivalent can be made with simpler models. It should be noted that simpler models will yield more general, less reliable results. Nonetheless, spatially-distributed estimates of SWE can be derived from point observations of SWE and remotely sensed areal extent of snow cover using techniques described below.

MODIFIED ELEVATIONAL DETRENDING

Linear precipitation-elevation relationships have been widely used to distribute precipitation in mountainous terrain (Chua and Bras, 1982, Phillips et al., 1992, Daly and Neilson, 1992). If cold-season precipitation varies with elevation in some reasonably consistent fashion, then it follows that snow water equivalent would exhibit similar tendencies.

The application of elevational detrending in our simple model focused on the reasonable representation of a local SWE - elevation relationship. Initial efforts involved only point observations of snow water. The results were not acceptable across the domain of the procedure, namely, the entire Western U.S. In some areas, the observational network was adequate to define the relationship, in most it was not. The key observational network deficiencies included (1) small SWE observation sample size within a (2) limited range of elevations.

The incorporation of satellite estimates of snow cover make elevation detrending possible in the Western U.S. Custom-generated composites of satellite areal extent of snow cover permit the identification of local snow line elevations which "tie down" the low end of the SWE - elevation relationship. Satellite-derived thematic raster images of snow cover are generated locally, on-demand, at the NOHRSC (Hartman, et al 1995).

The process of "modified" elevational detrending operates as follows for each grid cell in the area to be estimated.

1. Point observations of SWE within a user-defined search radius are identified and their average elevation and average snow value calculated (Figure 1).
2. From an areal extent of snow cover raster, the elevation of the nearest snow line is identified (Figure 2). If the snow line is too distant or is non-existent, a user-defined default snow line elevation is used in its stead.
3. A linear relationship between elevation and SWE is established from the snow line and the average elevation and the average SWE observation (sloped line in Figure 3).
4. Elevationally detrended residuals are computed for each SWE point observation using the relationship in (3) (blue lines in Figure 3).
5. An inverse distance weighted mean interpolated residual is computed (red line in Figure 3).
6. An estimate of snow value is computed using the relationship in (3) and the elevation of the pixel being estimated (yellow box in Figure 3).
7. The estimate in (6) is modified by the interpolated residual in (5) yielding the final estimated SWE (red box in Figure 3. Low and high resolution gridded estimates in Figures 4 and 5.)

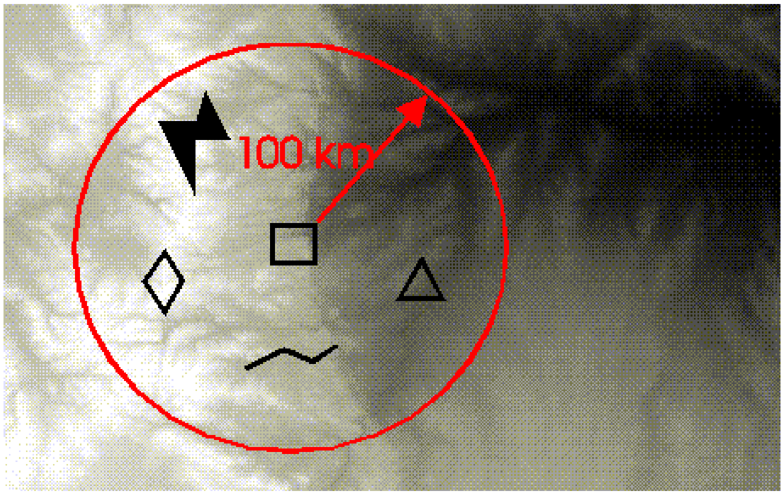


Figure 1: Snow telemetry (SNOTEL; lightning bolt), snow course (diamond), airborne gamma survey (line), and aerial marker observations within a specified distance of the estimated grid cell (square). The elevation of each observation and of the estimated grid cell is determined from the DEM which is displayed, in this figure, in varying shades of gray.

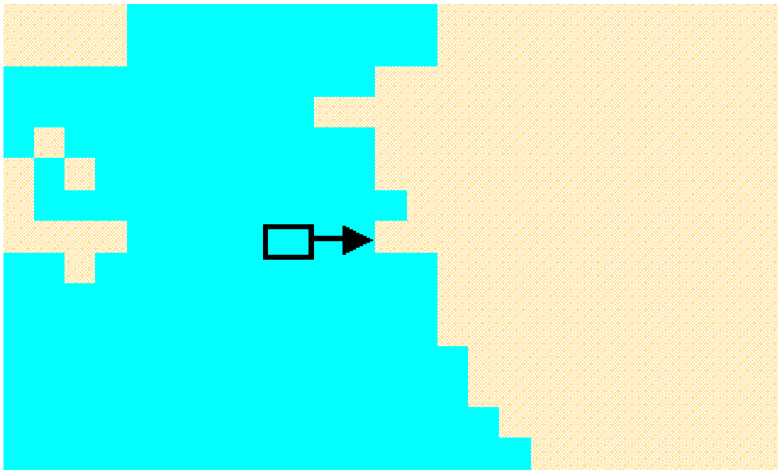


Figure 2: Areal extent of snow cover (blue) developed from satellite image processing. The arrow indicates the location of the snow line nearest the estimated grid cell. The snow line elevation is determined from the DEM.

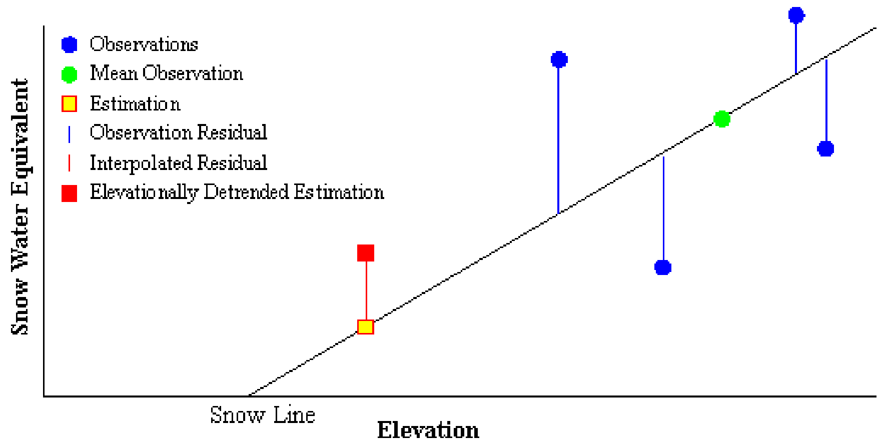


Figure 3: Snow value to elevation relationship calculated from the snow line and the mean observation.

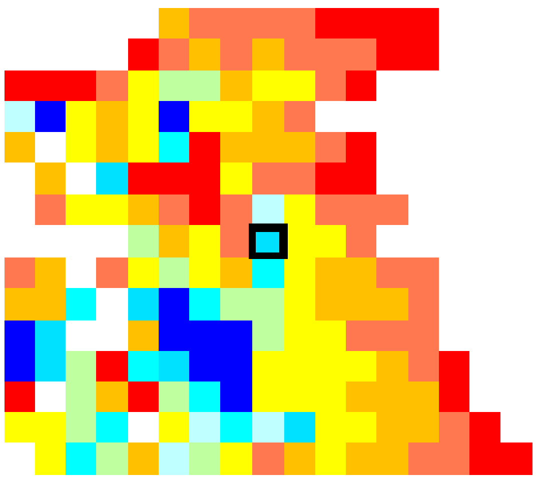


Figure 4: 300 arc second gridded snow values.

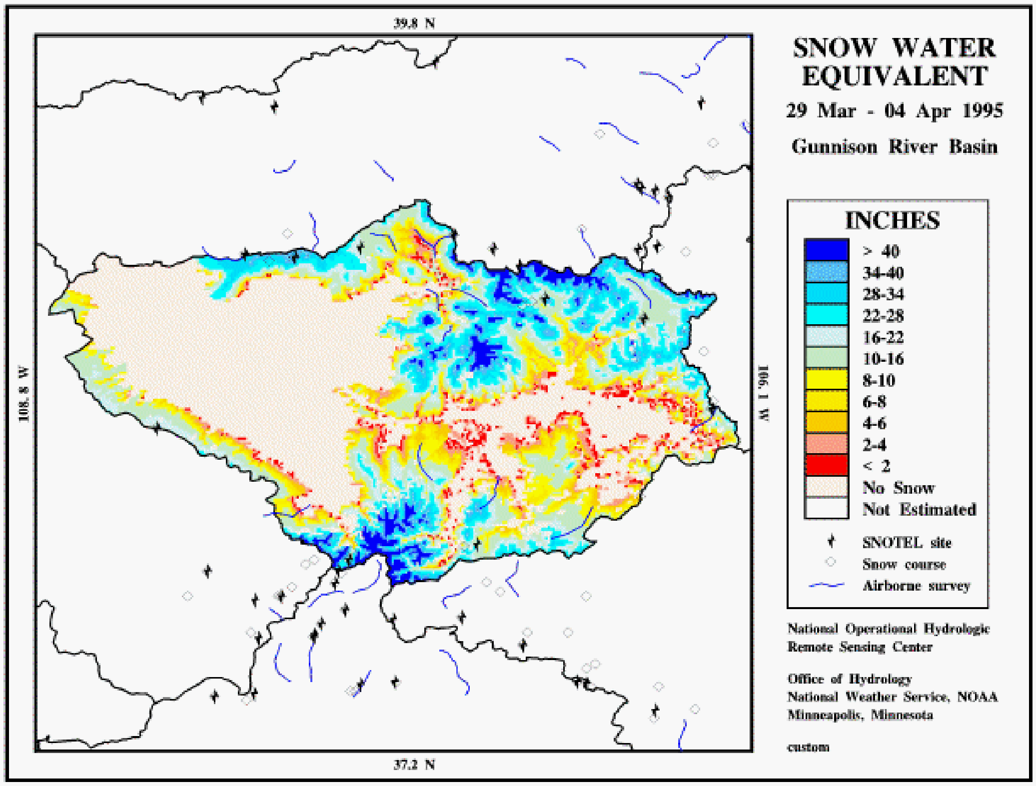


Figure 5: 30 arc second gridded snow values.

It is recognized that this simple modeling approach greatly generalizes the physical processes that result in the accumulation and melt of the mountainous snowpack. In truth, the observational network probably does not contain adequate information to accurately describe gridded snow water. The introduction of remotely sensed areal extent of snow cover data makes gridded estimation more feasible, but reason should be applied when making use of the results. Very little confidence should be placed on individual pixel values. However, when integrated within hydrologic basins, the results appear to be credible. Comparison of this approach with SEUS, an elaborate physical model of snow accumulation and melt, indicates that this simplified approach captures the spatial distribution and trends exhibited by the snowpack.

Estimation comparisons with SEUS in the Colorado Basin indicated the described procedure may overestimate at higher elevation. Work is underway at NOHRSC to "reshape" the snow water - elevation relationship above the elevation of the highest observation point.

CONCLUSION

A simple procedure for estimating gridded SWE using observational point values and satellite areal extent of snow cover has been developed for use in the Western U.S. The procedure was run throughout the winter of 1994-1995 and provided very credible results. A number of simple enhancements are planned. Over time, the operational estimates provided by the described model will be replaced by more physically-based procedures, such as SEUS.

Raster and image products developed by this procedure are available and updated at least once a week on the

NOHRSC HomePage (http://www.nohrsc.nws.gov).

REFERENCES

Chua, S., and R.L. Bras, "Optimal Estimators of Mean Areal Precipitation in Regions of Orographic Influence", Journal of Hydrology 57:23-48, 1982.

Daly, C., and R.P. Neilson, "Digital Topographic Approach to Modeling the Distribution of Precipitation in Mountainous Terrain", *In:* Interdisciplinary Approaches in Hydrology and Hydrogeology, American Institute of Hydrology, pp. 437-454, 1992.

Day, G.N., "A Methodology for Updating a Conceptual Snow Model With Snow Measurements", NOAA Technical Report NWS 43, National Weather Service, Silver Spring, Maryland, 1990.

Hartman, R.K., A.A. Rost, and D.M. Anderson, "Operational Processing of Multi-Source Snow Data", Proceedings of the Western Snow Conference, pp. 147-151, 1995.

McManamon, A., T.L. Szeliga, R.K. Hartman, G.N. Day, and T.R. Carroll, "Gridded Snow Water Equivalent Estimation Using Ground-Based and Airborne Snow Data", Proceedings of the Eastern Snow Conference, Quebec, Canada, pp. 75-81, 1993.

Phillips, D.L., J. Dolph, and D. Marks, "A Comparison of Geospatial Procedures for Spatial Analysis of Precipitation in Mountainous Terrain", Agricultural and Forest Meteorology, 58, pp. 119-141, 1992.

---

[1]*National Operational Hydrologic Remote Sensing Center*
Office of Hydrology, National Weather Service
1735 Lake Drive West
Chanhassen, Minnesota 55317-8582
Telephone (612)361-6610, Facsimile (612)361-6634
email: {rhartman, arost, danderson}@nohrsc.nws.gov
www: http://www.nohrsc.nws.gov

Robert K. Hartman[1], Andrew A. Rost[1], Donald M. Anderson[1]

Operational Processing of Multi-Source Snow Data

ABSTRACT

Near real-time quantitative assessments of areal extent of snow cover and snow water equivalent (SWE) provide valuable insight and information for water management, hydrologic forecasting, and emergency preparedness entities throughout the United States and Canada. Since 1986, the National Operational Hydrologic Remote Sensing Center (NOHRSC) of the National Weather Service (NWS) has provided these data in rasters and alphanumeric tabulations generated using satellite image processing, spatial interpolation, and GIS analysis.

Composite use of multi-source satellite (i.e., Advanced Very High Resolution Radiometer (AVHRR) and Geostationary Operational Environmental Satellite (GOES)), point (i.e., manual and telemetered observations), and line (i.e., airborne gamma radiation surveys) data is now accomplished through the newly developed NOHRSC Operational Product Processing System (OPPS). OPPS ingests and databases classified (i.e., snow, ground, cloud, and unknown) satellite areal extent of snow cover rasters and all identified point and line samples of SWE and snow depth. OPPS generative capabilities include user-prioritized composite raster generation and multi-source gridded estimates of SWE and snow depth. OPPS spatial estimation processes for SWE and snow depth combine point and line data with composited satellite areal extent of snow cover rasters. OPPS can 1) output rasters in any of several common GIS formats, 2) create displayable images with user-defined vector overlays, and 3) integrate raster data within basin boundaries to produce alphanumeric data products. OPPS products can be stored and distributed using the NOHRSC World Wide Web home page, FTP'd over Internet to interested cooperators, and sent to NWS offices over AFOS (Automation of Field Offices and Services).

INTRODUCTION

Near real-time quantitative assessments of areal extent of snow cover and gridded snow water equivalent (SWE) provide valuable insight and information for water management, hydrologic forecasting, and emergency preparedness agencies throughout the United States and Canada. Since 1986, the National Operational Hydrologic Remote Sensing Center (NOHRSC) of the National Weather Service (NWS) Office of Hydrology (OH) has provided these data in rasters and alphanumeric tabulations generated using satellite image processing, spatial interpolation and GIS analysis. Until recently, each of these processes were performed independently. They are now largely integrated in the newly developed NOHRSC Operational Product Processing System (OPPS).

Estimates of the extent, depth, and water equivalent of snow are developed at the NOHRSC from a variety of dynamic data sources including:

1. Satellite imagery acquired daily at a variety of resolutions from a variety of sensors (i.e., Advanced Very High Resolution Radiometer (AVHRR) and Geostationary Operational Environmental Satellite (GOES));
2. Airborne gamma radiation survey snow water equivalent data sets; and
3. Station observations reporting point data on snow water equivalent, snow depth, air temperature, etc. (i.e., SNOTEL, snow course, aerial markers, etc.).

Analyses of these data are supported by static data sets including hydrologic basin boundaries at various scales, thematic land use and land cover grids, and digital elevation models.

Prior to OPPS, each data type was analyzed by a separate data processing system. A given snow estimation product may have employed a variety of input data sources and therefore more than one process was required. For instance, to produce a tabulation of mean SWE by elevation band, the following processes were executed:

1. Acquire and analyze satellite imagery to produce an areal extent of snow cover map;
2. Spatially interpolate SWE station observations;
3. Mask the SWE spatial interpolation with the snow cover map to ensure that SWE is estimated only where snow is observed;
4. Stratify interpolated SWE by elevation bands; and
5. Average SWE within each elevation band for individual hydrologic basins.

While each individual process was highly automated, the integration of intermediate outputs into the final products was both labor intensive and procedurally inefficient.

OPPS DESIGN

OPPS was developed to meet the following design objectives:

1. To streamline, to the greatest extent possible, the production of snow estimation products in an operational environment;
2. To integrate, in an automated and objective manner, a wide variety of input data sources used to produce snow estimation products;
3. To develop and employ state-of-the-art spatial data processing algorithms tailored to the task of producing snow estimation products from integrated input data sources; and
4. To automate the dissemination of generated products.

A major objective in the design of OPPS was to automate data integration. Primarily through the use of spatial interpolation techniques and polygon membership modeling, OPPS is capable of integrating raster data with point, line and areal vector data. The integration of variable-resolution raster data is supported by run time data sub- and super-sampling functions allowing OPPS to define a range of output product resolutions without regard to the resolution of the input data.

The spatial integration of raster and vector data is supported by automated procedures which exploit the temporal distribution of the input data. Many OPPS processes are designed to evaluate data within windows of opportunity centered on a target date. Because OPPS is designed to address snow estimation on a continental scale, there is a strong possibility that suitable input data are not available for a given instant in time. For example, processes which require satellite image derived maps of areal extent of snow cover are often hampered by cloud cover. By integrating the cloud-free portions of multiple snow cover maps acquired during a window of opportunity, OPPS can minimize the impact that cloud cover has on the snow estimation process. Similar mechanisms were designed into OPPS for the treatment of each input data source.

OPPS consists of a series synchronized programs communicating with one another through an INFORMIX database server and system calls (Figure 1). The OPPS programs fall into three classes: Database, Analysis, and Product Development. The OPPS programs are supported by the OPPS database consisting of static and dynamic tabular and graphic data. The dynamic portion of the OPPS database is constantly updated by a wide variety of inputs. The remainder of this paper describes, in some detail, the OPPS components and their relationships in the production of OPPS outputs.

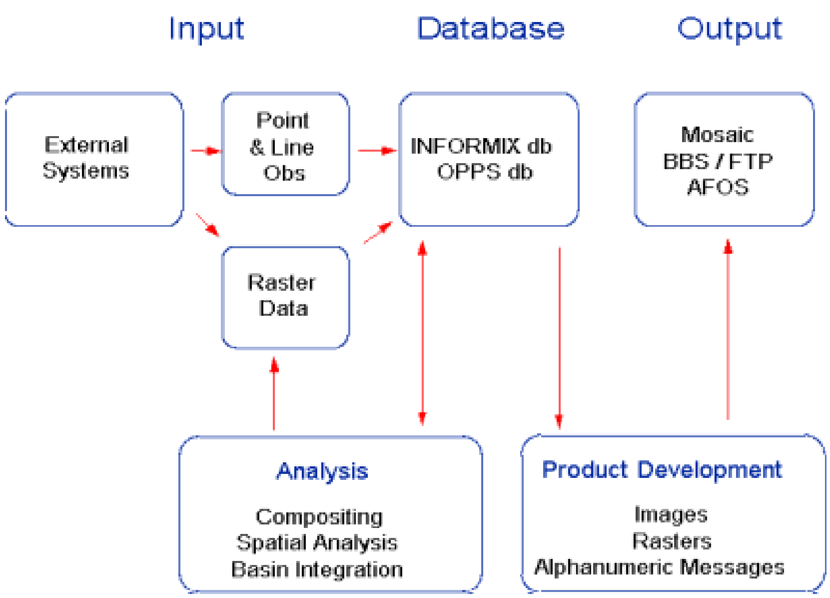


Figure 1. OPPS data/products/processing flow

OPPS INPUT

OPPS is designed to integrate data from a variety of sources, data-types, structures and formats. OPPS can handle both raster and vector data types. Raster data can be variable in resolution. Vector data may be either point, line, or area structures.

To minimize distortions associated with map-projected coordinates, OPPS requires that all of its inputs be in geodetic (longitude and latitude or Earth) coordinate pairs. All calculations are performed in the geodetic coordinate system. The World Geodetic System 1984 (WGS 84) horizontal datum and the National Geodetic Vertical Datum of 1929 (NGVD 29) were selected for OPPS on the basis of the availability of digital elevation model (DEM) data. Many of the analysis programs in OPPS model orographic processes and, as such, are highly dependent upon DEM data. The highest quality of DEM Data available in national coverage are in the WGS 84 and NGVD 29 datums.

Point observations of snow water, snow depth, precipitation, and surface air temperature are actively and passively acquired over Internet and are automatically ingested into the INFORMIX database. As airborne survey data become available, they also are ingested into the INFORMIX database.

Raster data inputs include areal extent of snow cover thematic rasters derived from the reclassification of satellite images acquired from the GOES and the NOAA polar orbiter satellites (AVHRR). Snow cover rasters are produced on a delay basis and are registered with the OPPS database as they become available.

OPPS is capable of ingesting raster data in GRASS, ARC/INFO, and Global Imaging formats. OPPS will be capable of ingesting raster data conforming to the SDTS raster profile in the near future.

Point observation data are registered and stored as INFORMIX database records. OPPS is capable of ingesting flat files into the INFORMIX database. All line and areal vector structures are stored in OPPS as individual binary flat files whose headers are stored in the INFORMIX database. Since many spatial data analysis systems (i.e., GIS) are capable of exporting flat files, the OPPS approach for dealing with these types of data allows a great deal of flexibility.

OPPS DATABASE

OPPS maintains two databases:

1. The OPPS database, a UNIX file system which organizes OPPS raster and vector files into subdirectories; and
2. The INFORMIX database which:
    1. Stores the characteristics, identity, and location of each OPPS-generated product;
    2. Stores the characteristics, identity, and location of each OPPS static data set (i.e., DEMS, basin boundaries, etc.); and
    3. Stores non-raster OPPS input data (i.e., station observations) and attributes.

In addition to storing information on products, the INFORMIX database is also used to control the flow of integrated OPPS processes. Certain OPPS products are the consequence of multiple data processing operations in which the output of one process is used as the input for a subsequent process.

These multiple processing operations are internally defined and, as such, are transparent to the user. When an OPPS product is requested, the request is registered in the INFORMIX database as a unique record. The individual processes required to produce that product are automatically queued by the OPPS subsystem. As subsequent processes are completed, their status is registered in the INFORMIX database and the next process in the queue is initialized. Queued processing continues until the final output product is generated.

OPPS ANALYSIS PROGRAMS

In its present state, OPPS consists of the following analysis programs:

Raster Compositing

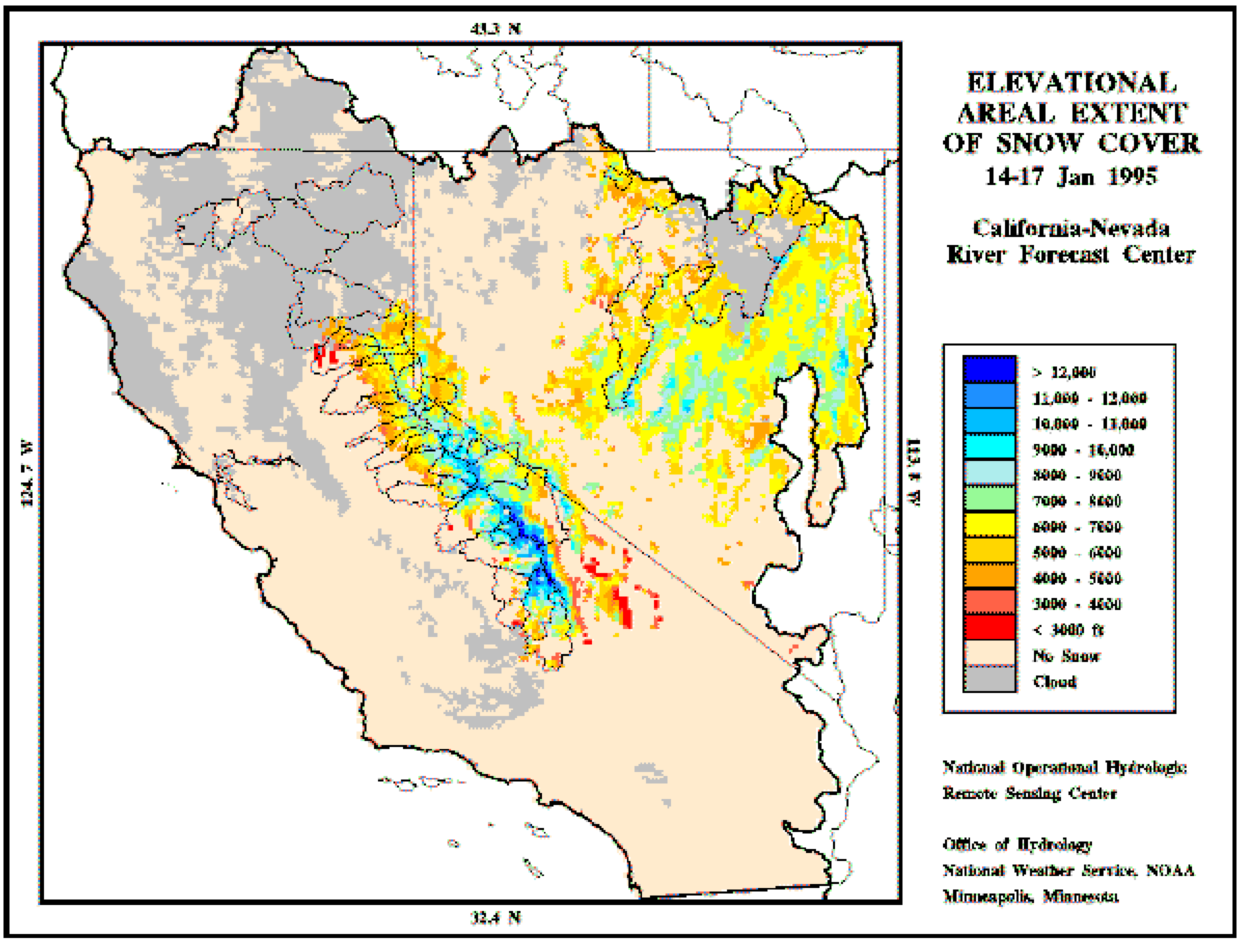
This program generates composites from existing rasters in the OPPS database. It has two functions:

1. To produce a single raster mosaic of a larger area from multiple rasters of smaller geographically distributed areas; and
2. To produce a single raster composite from temporally-distributed rasters occupying the same geographic area. As rasters are composited, unknown and cloudy pixels are replaced by pixels of known value.

Both functions can operate simultaneously and there is full control of how multiple rasters are integrated. Composite control is facilitated by prioritizing raster file attributes. Raster compositing may be constrained by vector feature outlines.

Areal Extent of Snow Cover by Elevation

This program produces a new raster in which the snow pixels in an areal extent of snow cover raster are represented by DEM elevation values. The resultant raster depicts the horizontal and elevational distribution of observed snow (Figure 2).



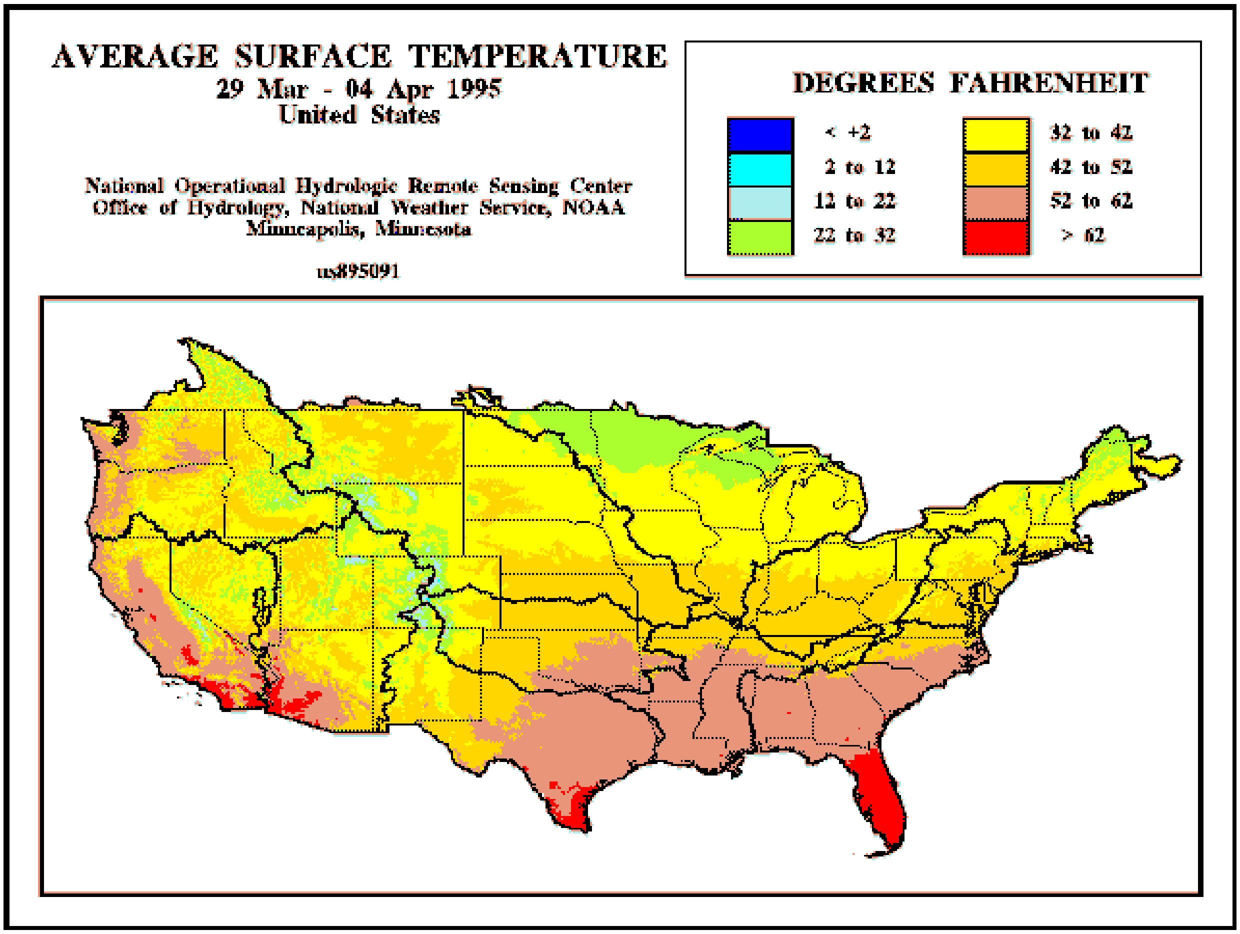
Inverse Distance Interpolation

This program accepts point and flight line observations as the x, y, and z sample data inputs into the inverse distance weighted mean spatial interpolation algorithm. The user controls where, geographically, sample points are to be gathered and how far the algorithm may search for sample observations.

No-snow pixels in an areal extent of snow cover raster can be used as sample observations of zero snow cover to constrain the interpolation. The snow cover raster may also be used to mask interpolated pixels for which no snow was observed by the satellite.

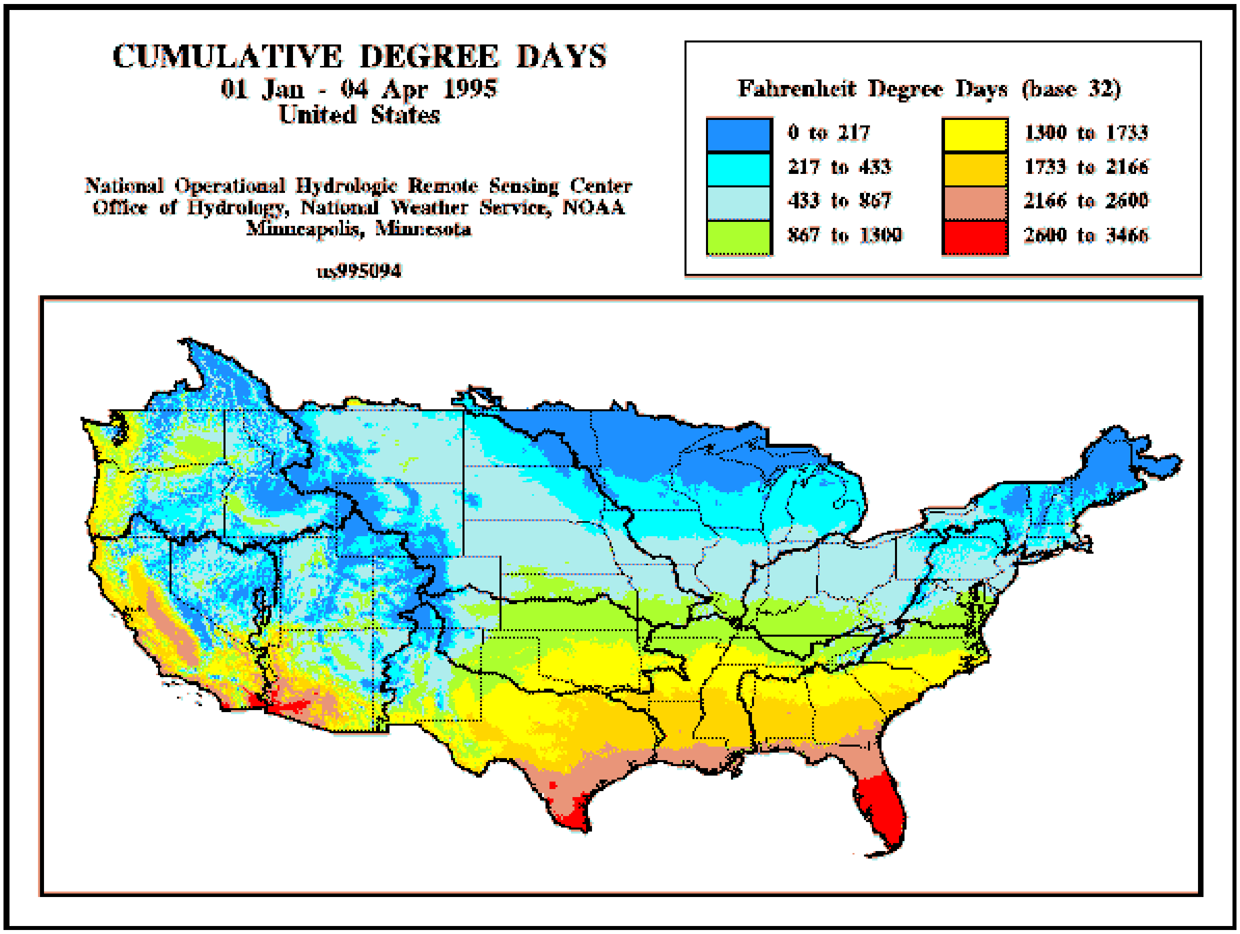
Temperature Interpolation

This program generates an inverse distance weighted mean interpolation of mean daily surface air temperature. Surface temperature observations are lapsed to a common elevation of 4500 meters. Temperatures interpolated at this elevation are lapsed back to the surface. The resultant product reflects the orographic effect on surface air temperature (Figure 3).



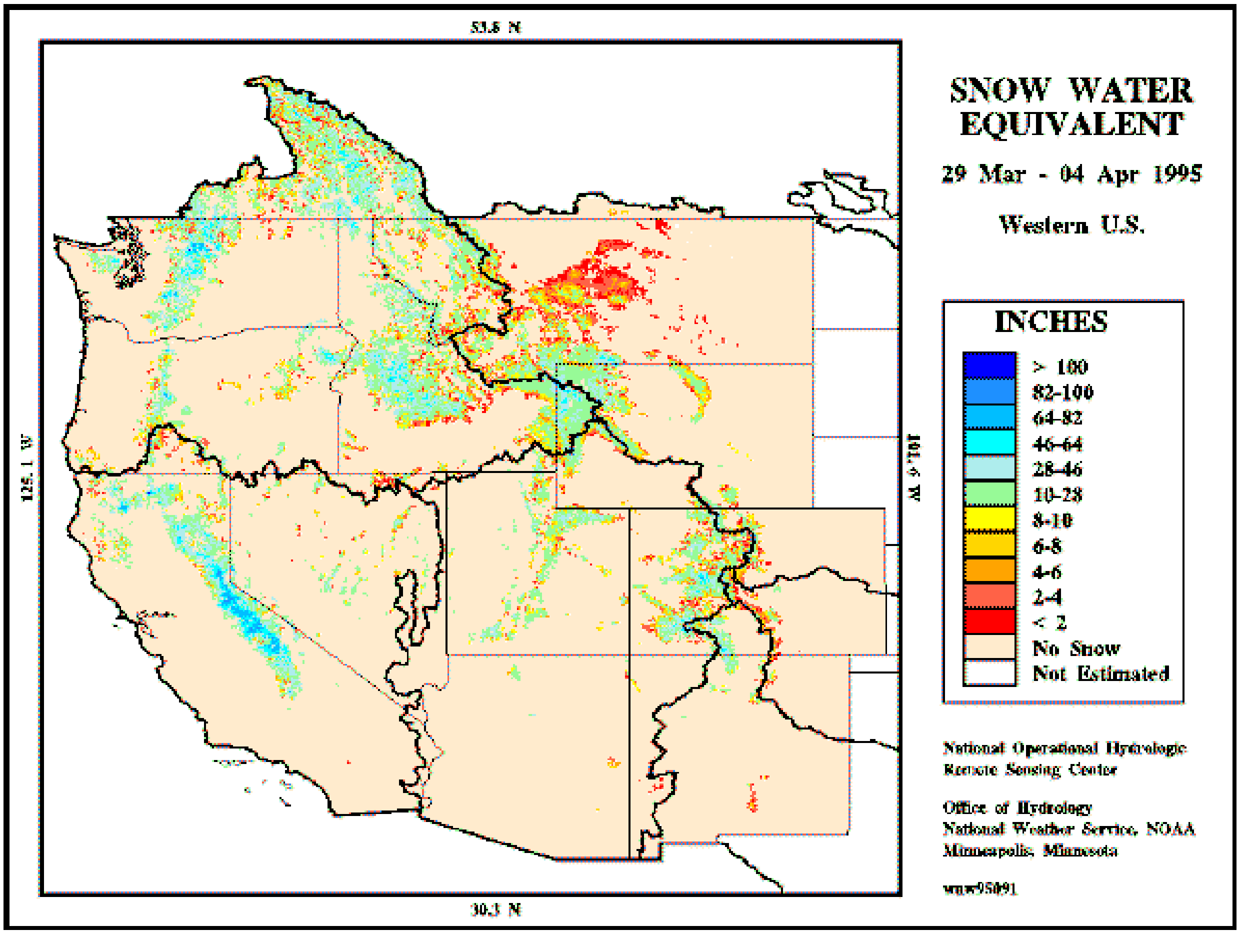
Degree Day Accumulation

This program sums, for each grid cell, the daily positive differences between interpolated surface air temperature and a specified base temperature. The output is useful for calculating the number of snowmelt degree days beyond a given date. The area calculated may be constrained by vector feature outlines (Figure 4).



Orographic Interpolation

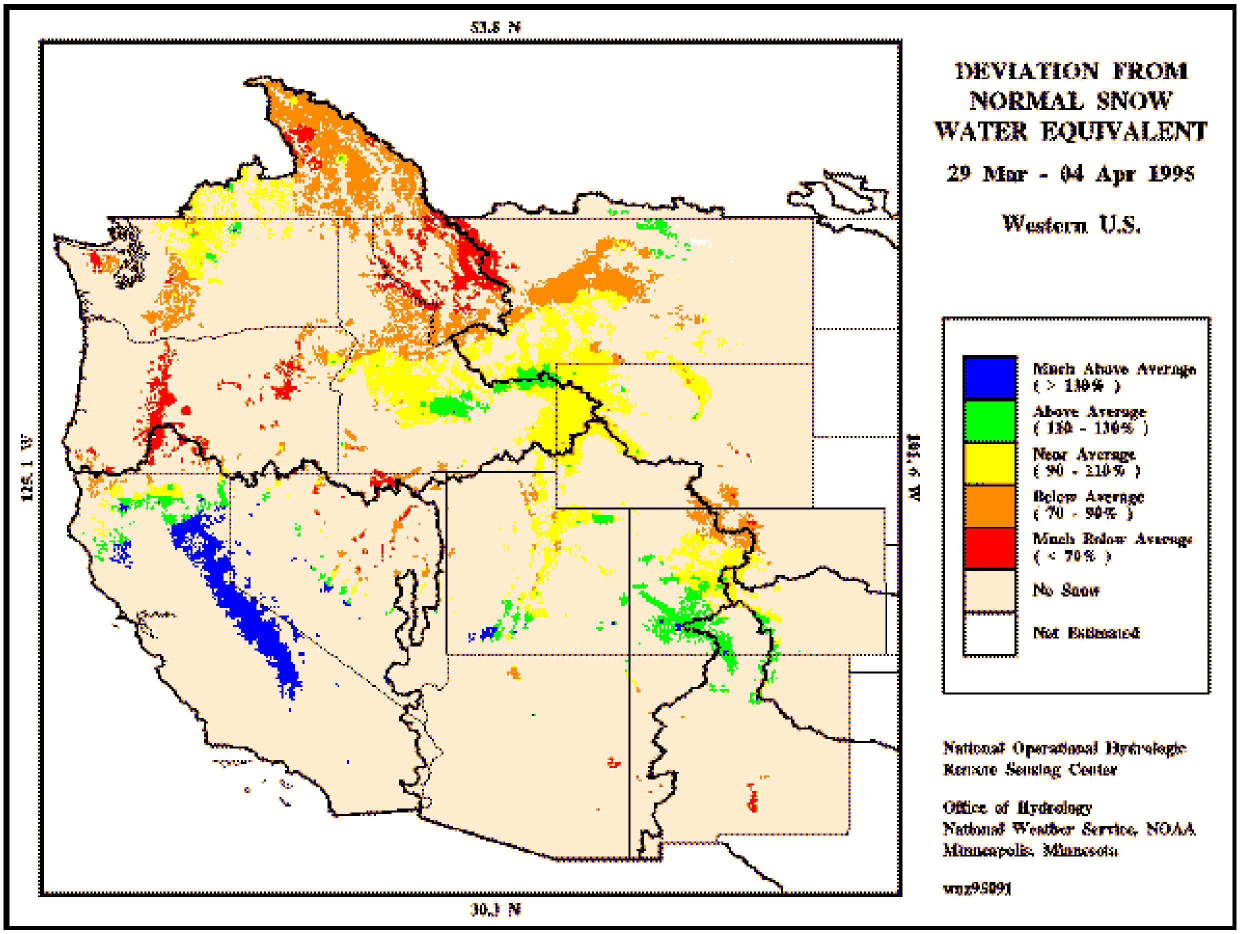
This program estimates SWE or snow depth in areas in which there are orographic dependencies (Figure 5).



For a given cell:

1. The nearest snow line elevation is computed on an areal extent of snow cover image; and
2. Ground-based SWE or snow depth point observations within a threshold distance are identified in the OPPS

A linear SWE (or snow depth) by elevation (orographic) relationship is established between the observed snow line and the mean point observation. A first approximation of SWE or snow depth for a given cell is calculated by applying the cell's elevation in this orographic relationship. The details of orographic interpolation are presented in an accompanying paper ("Spatial Distribution of Snow Water Equivalent Observations in Mountainous Terrain"). The SWE rasters generated by this program can be compared against historic data to determine deviations from normal SWE (Figure 6).



Basin Analysis

This program will, on a basin-by-basin analysis, compute mean areal snow values (i.e., SWE, snow depth, etc.) by predefined elevation bands. The resulting tabulations are stored in the INFORMIX database as time-stamped records. As basins are examined throughout the snow season, additional basin analysis records are related to each basin, yielding a time series of basin snow data.

OPPS PRODUCT DEVELOPMENT

OPPS generates a variety of gridded products derived from raster and vector inputs using processes designed specifically for snow distribution estimation. Basin-integrated values can be summarized into coded Standard Hydrometeorological Exchange Format (SHEF) alphanumeric products specifically tailored to a specific user's needs.

All of the gridded products generated by OPPS reside in the OPPS database in a binary format designed to accommodate the needs of the NOHRSC. OPPS is capable of exporting these rasters into a variety of formats including GRASS, ARC/INFO ASCII grids, and Global Imaging (essentially a flat binary raster with an 80 byte header).

While these formats respond to the needs of those interested in performing further data analysis, some end users are interested primarily with images (pictures). To meet those needs, OPPS can also generate GIF, TIFF, McPaint, and PCX formatted image files. A variety of vector overlay features may be incorporated into these images, including states, basins, and observation points.

OPPS OUTPUT

The NOHRSC data are used operationally by the NWS, the Army Corps of Engineers and other Federal, state, and private agencies when issuing spring flood outlooks, water supply outlooks, river and flood forecasts, and reservoir inflow forecasts. In an effort to simplify the distribution of products to regular data users, OPPS has incorporated a sophisticated data distribution subsystem which automatically routes relevant and/or desired OPPS output products to agencies who have registered their Internet FTP addresses with the NOHRSC. Users can also access OPPS products via the NOHRSC World Wide Web Home Page (http://www.nohrsc.nws.gov). Alphanumeric products are sent to NWS offices over AFOS (Automation of Field Offices and Services).

CONCLUSION

The NOHRSC is an operational center within the National Weather Service Office of Hydrology which has, as one of its major responsibilities, the task of mapping snow cover characteristics for all of the United States and portions of Canada. While the NOHRSC has always employed state-of-the-art techniques in accomplishing this task, OPPS, for the first time, allows the snow hydrologist access to the full value of the available input data.

Through the integration of a broad variety of input data sources, the exploitation of temporal processes, and the design and implementation of data management and analysis programs tailored to snow distribution analysis, OPPS is capable of providing and distributing near real time analysis of the highest quality currently available.

---

[1]*National Operational Hydrologic Remote Sensing Center*
Office of Hydrology, National Weather Service
1735 Lake Drive West
Chanhassen, Minnesota 55317-8582
Telephone (612)361-6610, Facsimile (612)361-6634
email: {rhartman, arost, danderson}@nohrsc.nws.gov

Mirror-Image Round Robin Spatial Data Partitioning : a Case Study With Parallel SEUS

A. A. El Haddi[1,2]
S. Shekhar[2]
R. Hills[1]
A. McManamon[1]

Fri Dec 15 10:56:58 CST 1995

Click here for the postscript version of this paper.

Abstract:

The National Weather Service interpolates snow conditions over numerous hydrologic basins to obtain snow water equivalent estimates and associated errors for gridded fields with 1 Km resolution. Solving problems of this scale involves an enormous number of computations and data input and output operations. Using sequential implementation to obtain gridded estimates may require execution times ranging from hours to days. Thus it becomes infeasible to solve large problems of national scale interactively or in a time frame suitable for near real time forecasts and flood warnings.

In this paper we discuss some techniques used to speed-up the computation by partitioning physical space into sub-domains. Spatial operations relating to each sub-domain are executed over a distributed heterogeneous memory architecture. Results from various strategies for distributing computations are discussed. A performance analysis is presented for the parallel implementation of snow estimation and updating system of the National Weather Service. A new Mirror-Image Round-Robin(MIRR) data partition technique is introduced. Many of these techniques have applications in developing high performance and distributed geographic information systems(DGIS).

1 Introduction

You are watching your favorite TV program when an unexpected voice interrupts with the following message ``The National Weather Service(NWS) has issued a severe flood warning for Fort Wayne. A crest of 9.50 feet above flood stage for the Maumee River at Anthony Boulevard is expected. The Red Cross, Civil Defense, Lutheran Services and the Salvation Army are preparing for major ..., please go to the nearest ...''. This could have been a real scenario, according to a study of flood damages at Fort Wayne, Indiana [Carroll and Marshall, 1985].

NWS maintains a set of applications to estimate snow-water equivalent (SWE) that are used to issue water forecasts and flood warnings for the country. SWE is used by NWS hydrologists to quantify the stream flow and to forecast water supplies for the United States and to manage this resource for various competing needs (e.g. domestic use, irrigation, hydro-power, etc). Government and state agencies rely on these forecasts to prepare for flood disasters and issue early flood warnings.

NWS applications use multi-source data to interpolate snow-water equivalent over large areas [Carroll, 1995][Hartman et al., 1995]. These data are obtained by satellites, low flying aircraft, ground based sensors and sampling crews. A spatial model is used to obtain a gridded SWE over larger areas and to obtain areal snow-water equivalent estimates over a river basin. Each areal estimate has an associated uncertainty, expressed in terms of its associated standard deviation. The accuracy with which areal snow-water equivalent is estimated is critical since lives are involved, economic losses due to floods are enormous and the need to manage water resources for the competing needs of irrigation, domestic use and hydro-power are crucial. It is also imperative that the snow-water equivalent estimation is done accurately in near real time. The National Operational Hydrologic Remote Sensing Center (NOHRSC) maintains a Snow Estimation and Updating System (SEUS) that is used to compute gridded SWE. SEUS is a software system developed by NWS to operate on the Geographical Resources Analysis Support System (GRASS). Readers can find more details on SEUS and snow estimation problems in [Fread et al., 1995][McManamon et al., 1995][McManamon et al., 1993][Day, 1990][Carroll et al., 1995].

In this paper we will focus on the spatial computational aspects of SEUS and propose a new mirror-image round-robin data partition technique and a next generation Parallel SEUS (PSEUS). Section 2, describes the application domain and the data types used by SEUS and the spatial interpolation methodology. Section 3 describes the computational problems associated with gridded snow estimation using SEUS. Section 4 describes data partitioning, a parallel version of SEUS (PSEUS) and an experimental evaluation. Section 5 is a summary and discussion of current and future research.

2 Application Domain: Snow Estimation

The NWS uses a spatial prediction model to derive gridded snow-water equivalents over many basins in the United States. The model relies on spatial correlation among the data and geostatistical techniques to estimate SWE where no observed data exists. A detailed review of the SEUS methodology can be found in [Day, 1990]. The gridded snow-water equivalent is used as input to a snow ablation and accumulation model that uses observed temperature and precipitation to simulate snow cover conditions [Anderson, 1973]. [McManamon et al., 1993] and [Carroll et al., 1995] describe the use of ground and airborne data to compute gridded estimates of snow-water equivalent and associated standard deviates.

2.1 Data types

From a modeling perspective, the data used to interpolate gridded snow-water equivalent falls into two categories: point and line data. From a computational point of view however, the data can be classified into three types: Areal, point and line data.

2.1.1 Areal data

The NWS divides the coterminous US into 12 hydrologic regions served by NWS River Forecast Centers(RFC). Each RFC region is sub-divided into as many as 800 hydrological basins.

Each basin in this study has a certain number of ground data collection points and airborne snow survey flight lines. Figure 1 depicts the relationship between all data types. The Colorado Basin

RFC, for example, has over 200 NWS forecast basins. We will use notation $R_i(i = 0, ..., r-1)$ to represent RFCs, where r is the total number of regions. Basins will be represented using $A_k(k = 0, ..., nb-1)$, where nb is the total number of basins in region i.

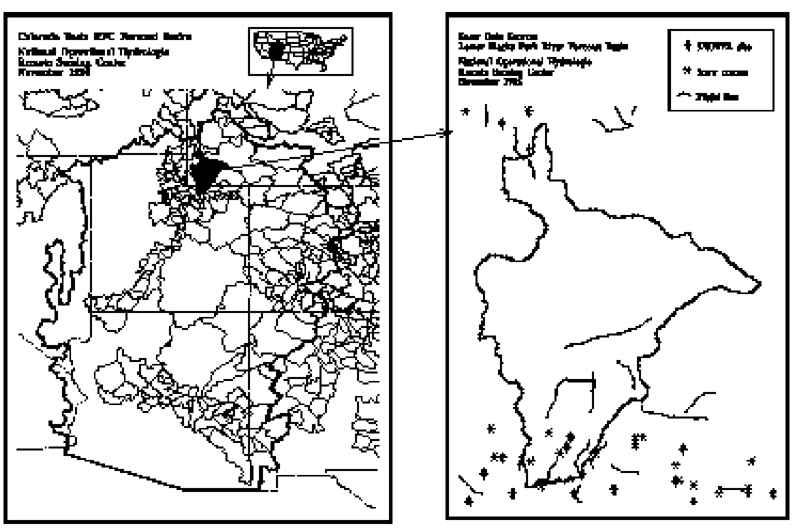
2.1.2 Line data

Line data are SWE observations made from a network of over 1800 flight lines. NOHRSC uses low-flying aircraft (operated by NOAA's commissioned Officer Corps) to measure natural terrestrial gamma radiation before the snow season and during the snow season. The technique is based on $^{40}K, ^{238}U, ^{208}Tl$ isotopes radiation attenuation due to the water mass in the snow cover. Each flight line is approximately 16 Km long, 300 m wide. Each flight line is subdivided into one or more segments. Details of the gamma radiation technique can be found in [Carroll et al., 1985][Carroll and Vadnais, 1980][Peck et al., 1980][Fritzsche, 1982].

The technique used to estimate SWE for each basin A_k uses flight lines $F_i(i = 0, ..., nfl-1)$ where nfl is the total number of flight lines that are within the basin itself or within adjacent basins.

2.1.3 Point data

Point data are SWE observations made from a network of ground point data composed of snow course and automated snow data telemetry (SNOTEL) sites. [McManamon et al., 1993], reported that the Natural Resources Conservation Service (NRCS) collects these ground data at over 2000 locations across the west. Snow course data are obtained by sampling snow and measuring its water content. We will use $G_i(i = 0, ..., ng-1)$ to denote the ground points in each basin A_k, with ng being the total number of ground points used in a basin's SWE computation. Each basin A_k is divided into a grid of points $GR_i(i = 0, ..., npt-1)$ where npt is the total number of points in the grid. These are the points for which snow-water equivalent is to be computed. They will be referred to as grid points. The number of grid points in each basin are proportional to the basin size.



(a) (b)

Figure 1: US map and relationship between data types (a) Basins within one RFC (b) Ground and Flight line data(Courtesy Don Anderson, NWS, NOHRSC).

2.2 SWE Interpolation

Snow-water equivalent is estimated using simple kriging. [Cressie, 1991] describes ordinary kriging, or optimal prediction, as making inferences on unobserved values of a random process $Z(i)$. These values are modeled using:

$$Z(s_i) = (Y(s_i) - \mu(s_i))/\sigma(s_i)$$

$Y(i)$ is the non-standardized value and has a non-zero mean. $Z(s_i)$ is the standardized SWE for both ground point and flight line data. Z has a mean of 0.

Ground and airborne data are used to compute the best linear predictor of $Z(s_0)$:

$$\hat{Z}(s_0) = \sum_{i=0}^{npt-1} \lambda_i Z(s_i)$$

Where:

λ_i are the simple kriging coefficients estimates. The vector λ is computed as follows:

$$\lambda = \sum^{-1} C_{s_0}$$

where

$$C_{s_0} = (cov(Z(s_0), Z(s_0)), \ldots, cov(Z(s_0), Z(s_{npt-1}))$$

$\sum$ is the $n \times n$ matrix whose $(i,j)^{th}$ element is $cov(Z(s_i), Z(s_j))$.

For a basin A_k whose grid points are $GR_i, i = 0, ..., npt - 1$ the standard deviate $\hat{\sigma}_k$ is modeled using

$$\hat{\sigma}_k = \frac{\sum_{i=0}^{npt-1} \sum_{j=0}^{npt-1} \hat{\sigma}(s_i)\hat{\sigma}(s_j)(cov(Z(s_i), Z(s_j)) - C_{s_i} \sum^{-1} C_{s_j})}{(npt)^2}$$

where $\hat{\sigma}(s_k)$

is the estimate of the standard deviation for a grid point s_k. More details about the interpolation can be found in [Carroll et al., 1995]

3 Serial Implementation of SEUS: SSEUS

The computation of the estimated SWE at each grid point can be broken down into seven steps, S1 through S7 shown in figure 2. It also computes the total basin SWE and associated uncertainty for A_k

each basin . These steps are representative of serial implementation such as serial SEUS(SSEUS).

S0: do S1, S2, S3, S4, S5, S6, S7 where:
S1: Input [S1.1] input ground points data. [S1.2] input flight line data. [S1.3] input grid points data.
S2: Standardize Means [S2.1] standardize ground point means. [S2.2] standardize flight line means. [S2.3] standardize grid point means.
S3: Compute covariance Matrix (Σ) [S3.1] compute Cov(Flight lines , Flight lines). [S3.2] compute Cov(Flight lines , Ground points). [S3.3] compute Cov(Ground points , Ground points).
S4: Compute inverse of covariance matrix(Σ 1)
S5: Compute Covariance Matrix with Grid points. [S5.1] Compute Cov(Flight Lines, Grid points). [S5.2] Compute Cov(Ground points, Grid points).
S6: Compute Gridded and total SWE estimates.
S7: Compute Total standard deviation

Figure 2: Pseudo algorithm to compute a single Basin (A_i) SWE and total variance.

To determine the run time of SEUS model as the characteristics of its instances change, we have used profiling, analysis, and measurement. Table 1 shows the number of iterations for each of the modules S1 through S7. We have analyzed the code and expressed the worst case number of iterations as a function of the number of ground points (ng), the total number of flight lines (nf), the total number of segments per flight line (ns_i) and the total number of grid points (npt) for each basin A_i.

Module	Worst case	Average (n)	Average (npt)
S1	$g(\sum_{i=1}^{nf} ns_i + ng + npt)$	$O(n^2)$	$O(npt)$
S2	$g(\sum_{i=1}^{nf} ns_i + ng + npt)$	$O(n^2)$	$O(npt)$
S3	$g(\frac{1}{2}\sum_{i=1}^{nf} ns_i^{\ 2} + ng\sum_{i=1}^{nf} + ng^2)$	$O(n^4)$	$O(1)$
S4	$g(\sum_{i=1}^{nf} ns_i + ng + npt)$	$O(n^2)$	$O(npt)$
S5	$g(npt\sum_{i=1}^{nf} ns_i + (ng)(npt))$	$O(n^3)$	$O(npt)$
S6	$g(npt(nf + ng))$	$O(n^2)$	$O(npt)$
S7	$g(\frac{1}{2}((npt)^2(ng + nf)^5)$	$O(n^7)$	$O(npt^2)$

Table 1: Serial SEUS worst and average time complexities for each module. Worst case is expressed as a function of actual variable ranges. Average case is expressed as a function of n which is a combination of all variable ranges for a given basin A_i or a region R_i, $n \equiv min(npt, ns_i, nf, ng)$.

The time complexity shows the run time as a function of the number of ground points(ng), the number of flight lines (nf), the number of segments per flight line (ns) and the number of grid points (npt) for which snow-water equivalent is to be computed. In our set-up, npt is always very large compared to nf, ns_i and ng. For example, the small Animas Basin used by Carroll et al(1990) had 2681 grid points, 14 ground points and 4 flight lines. The number of segments per flight line were 47, 62, 65 and 61 for flight lines 1, 2, 3 and 4 respectively. The number of grid points is proportional to the basin surface size. The relation $npt \gg ns_i \gg ng \gg nf$ holds for all basins in all regions. Table 1 shows the time complexity only for a single basin in one RFC. Using the same algorithm for multiple basins and multiple regions, the time complexity becomes a $O(npt^3)$ problem.

Serial SEUS executable was profiled using gprof on an HP 735-99Mhz running HP-UX 9.05. Execution profiles enabled us to see the number of calls to each module and its descendant functions. It also showed an approximation of the execution time for each module and descendants. A partial listing of the results where nf and ng were held constant is shown in table 2. The results suggest that computing the total variance represents the largest portion of the execution time. Our analysis of the time complexity concur with the results of profiling.

Examining the function that computes the total standard deviates(S7 is implemented as **compute_area_std_deviate()** as shown in figure 3), we found that we have a time complexity largely dependant on the total number of grid points. Our experimental data show that this is actually a $O((npt)^2)$ problem, since ng and nf are very small compared to npt. We can take advantage of the symmetry and loop only on $\frac{1}{2}npt^2$. Also Since Σ^{-1} is a constant for any combination i, j, we can pre-compute $C_j \Sigma^{-1}$ or $C_i \Sigma^{-1}$ before starting to loop on i or j.

	npt=100		npt=200		npt=1000		npt=2681	
Module	T(sec)	%Total	T(sec)	%Total	T(sec)	%Total	T(sec)	%Total
S1	0.186	16.87	0.189	6.23	0.203	0.33	0.200	0.05
S2	0.000	0.00	0.000	0.00	0.000	0.00	0.000	0.00
S3	0.186	16.87	0.192	06.32	0.185	0.30	0.179	0.04
S4	0.003	0.27	0.004	0.12	0.002	0.00	0.002	0.00
S5	0.121	10.93	0.231	7.59	1.166	1.87	3.160	0.72
S6	0.011	00.95	0.020	0.66	0.123	0.20	0.319	0.07
S7	0.590	53.47	2.396	78.92	60.705	97.31	436.171	99.12
Total	1.102	100.00	3.036	100.00	62.385	100.00	440.039	100.00

Table 2: Results of gprof on the serial version of SEUS for small grid sizes (times are in seconds). Each value is a mean of 20 repetitions. A 300 second sleep time follows each run to take into account different system loads.

```
// Compute σ̂ = sumsig .
double compute_area_std_deviate(double c, double *plat, double *plon,
       double *sig, double **sigmainv, double *C_i, double *C_j)
{
double sumsig,rho,dist;
int i,j;
sumsig=0
     for ( j=0; j < npt; j++ ) {
        // mul_mat_vec() is a matrix × vector routine.
        lambda = mul_mat_vec(sigmainv,Cj);          // λ = C_j Σ^-1
        for ( i= j+1; i < npt ; i++ ) {
           // Compute Distance between two points i with a latitude plat[i]
           // and a longitude plon[i] and a point j with latitude plat[j]
           // and a longitude plon[j].
           dist = distance( plat[j],plon[j], plat[i], plon[i]);
          // Compute the modeled cov(Z(s_i),Z(s_j))
           rho = c*exp(d*dist);
           if ( rho == c ) rho=1;
           // mul_vec_vec() is a vector × vector routine.
           sumsig = sumsig + sig(i)*sig(j)*(rho - mul_vec_vec(lambda,C_i));
        }
     }
     sumsig = 2*sumsig/n*n;
return sumsig;
}
```

Figure 3: Serial Version of stub computing total variance for a basin.

4 Parallel implementation: PSEUS

In section 3, we have shown that the module S7, which is implemented by the **compute_area_std_deviate()** function, is the slowest of all modules. This function computed $\hat{\sigma}$ for each basin. The function was slightly modified to run on a network of workstations, using a master/slave model and the Parallel Virtual Machine libraries [Geist et al., 1993]

This model is used to execute **compute_area_std_deviate()** in a distributed scheme on a network of workstations. In this model, one host was designated as the master, while the rest of the hosts were designated as slaves executing tasks on behalf of the master. Figure 5 depicts the data flow among workstations. The algorithm used by the master consists of the serial modules S1, S2, S3, S4, S5, S6, S8, S9. S1 through S6 are as explained in the serial implementation of gridded SWE. S8 and S9 are shown in figure 4.

S8: Send the data to all hosts
- [S8.1] Configure the virtual machine
- [S8.2] Start all slaves with **compute_area_std_deviate()**
- [S8.3] Broadcast all data required to compute $\hat{\sigma}$

S9: Receive partial results from each host
- [S9.1] Wait until all results are in
- [S9.2] Reassemble $\hat{\sigma}$ from all hosts

Figure 4: Master modules involved in parallelism.

Modules S1 through S6 output is used as input for module S7. Once the master is done executing modules S1 through S6, it farms out S7 to the slaves. Each individual slave determines the physical domain that will constitute its computation domain and runs S7. The master and the slaves constitute a virtual machine. At any one time, all slaves run exactly the same copy of the code (S7). This task has an identifier that we will use interchangeably with processor identifier. The task identifier for processor i will be represented using P_i.

Our goal is to compute $\hat{\sigma}$ as shown in section 3. We will decompose $\hat{\sigma}$ as follows:

$$\hat{\sigma} = \frac{2\sum_{P_i=0}^{np-1} \hat{\Psi}(P_i)}{npt^2}$$

where: $\hat{\Psi}(P_i)$ is the non weighted portion of $\hat{\sigma}$ computed by processor P_i and np is the total number of processors used to compute $\hat{\sigma}$. $\hat{\Psi}(P_i)$ can be further refined to be expressed in terms of the domain over which it is computed. The grid is partitioned row-wise to compute

$$\hat{\Psi}(P_i) = \sum_{k=row_i}^{row_n} \sum_{j=k+1}^{npt-1} \hat{\sigma}(s_k)\hat{\sigma}(s_j)(cov(Z(s_i), Z(s_j)) - C_{s_k} \sum^{-1} C_{s_j})$$

Rows row_i and row_n are the lower and upper bounds of the row values from the grid. These are not necessarily consecutive as we will show.

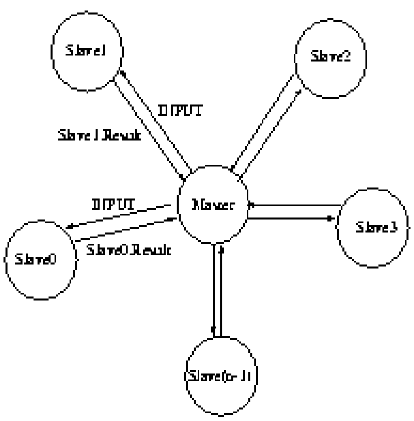


Figure 5: Master/Slave PSEUS model and data flow

Once each processor receives its data from the master, it computes $\hat{\Psi}(P_i)$ only over the portion of the grid which constitutes its partition domain. Three strategies were used to partition data among processors. These strategies are:(1) Contiguous Row Blocking(CRB), (2) Round-Robin(RR) and (3) Mirror-Image Round-Robin(MIRR). The reader can find a survey of other partitioning techniques for GIS data in [Shekhar et al., 1995][Shekhar and Liu, 1995]. In all of our strategies, the basin is represented as an $\mathrm{npt} \times npt$ grid space.

```
// Computes bounds for consecutive rows assigned to a processor.
getrows1(int Pi, int npt, int np, int *lower_bound, int *upper_bound )
{
     // compute upper and lower bound range of rows
     // allocated to each processor Pi
     if ( npt > 0 && Pi< np ) {
       *lower_bound = Pi*(npt/np) ;
       if ( Pi< np - 1 )
            *upper_bound = *lower_bound + npt/np -1;
       else *upper_bound = npt - 1; // last processor gets the rest
     }
}
```

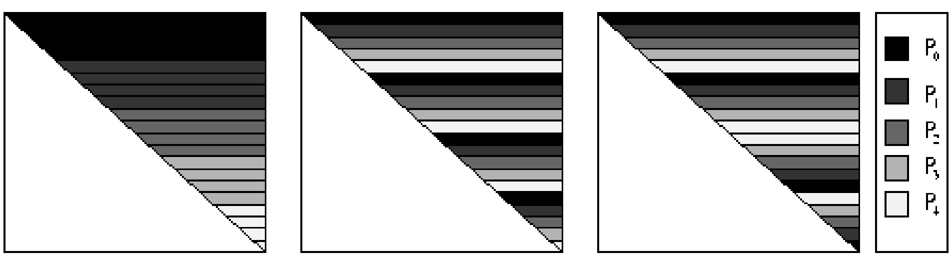
Figure 6: Contiguous Row Blocking partition strategy: allocates equal rows to all processors

In CRB strategy , we attempt to allocate an equal number of consecutive rows to each processor using the stub of figure 6. Using this scheme, each processor $P_i\{i = 0, ..., np - 1\}$ is allocated $(npt \div np)$ rows (where $\div$ is the integer division operator). Processor, P_{np-1}, is allocated an additional $npt - (np - 1)(npt \div np)$ rows. Figure 7(a) illustrates this partition method. In this strategy, each processor i is allocated rows $R_j, R_j \in [i(npt \div np), (npt \div np)(i+1) - 1]$. The number of cells allocated to processor i is given by:

$$P_i c = \frac{((npt \div np)(2npt - (2i+1)(npt \div np) - 1))}{2}$$

The total number of cells processed by the virtual machine is given by:

$$T_c = \sum_{i=0}^{np-1} P_i c = \frac{npt(npt-1)}{2}$$



(a) (b) (c)

Figure 7: Partition of a 20 x 20 grid among 5 processors.(a) Contiguous Row Blocking (b) Round-Robin (c) Mirror-Image Round-Robin

We use a grid size of $20\ npt \times 20\ npt$ to illustrate this partition strategy. We obtain the results shown in table 3. The results are also expressed as a proportion of the total number of cells processed.

	CRB strategy		RR strategy		MIRR strategy	
Processor	# of Cells	% Total	# of Cells	% Total	# of Cells	% Total
P_0	70	36.84	46	24.21	38	20.0
P_1	54	28.42	42	22.11	38	20.0
P_2	38	20.00	38	20.00	38	20.0
P_3	22	11.58	34	17.89	38	20.0
P_4	6	3.16	30	15.79	38	20.0

Table 3: Total number of cells and proportion allocated to each slave *(npt=20, np=4, T_c=190)*)

We note that this allocation scheme leads to a load imbalance where processors do not get equal work. Processors with a low identifier have a high number of cells allocated to them. In this example, when the virtual machine is configured with 5 hosts, P_0 processes 37 percent of the total cells while P_2 processes only 3 percent. Thus a few processors do most of the computations while the others spend most of the time waiting for processors with a larger cell allocation to finish computing their portion of $\hat{\sigma}$.

To distribute the load among all slaves, CRB partition strategy was refined to obtain Round-Robin partition strategy. Figure 7(b) shows how the grid space is partitioned using the algorithm of figure 8. In this strategy, each processor P_i is allocated rows $\mathrm{i} + \mathrm{j(np)}, \mathrm{j}=0,1,\ldots, (\mathrm{npt} \div np) - 1$. Each of the remaining rows *(r= npt **modulus** np)* are allocated to processors 0 to $\mathbf{r\text{-}1}$. Thus, each processor P_i is allocated a total number of cells as follows:

$$P_i c = d(npt - 1 - i) - \frac{np(d-1)d}{2} + \alpha(npt - 1 - i - d(np))$$

where :

$\mathrm{i} = 0, \ldots, \mathrm{np\text{-}1}$

$\mathrm{d} = \mathrm{npt} \div np$

$\alpha = 1, \forall i < r$

$\alpha = 0, \forall i \geq r$

```
// Round—Robin Partition strategy:
// Given npt grid points and np processors:
// Get rows assigned for processor Pi.
// Rows are stored in the array rows.
getrows2(int Pi,int npt, int np,int *rows)
{
     int d,r,j;
     if (np == 0 ) np = 1;
     d= (npt) / np ;
     r= (npt) % np ;
     for ( j=0; j < d; j++ ) {
       rows[j] = Pi + j * np;
     }
     if ( r > 0 && Pi < r)
     rows[j++] = Pi + j*np ;
     rows[j]=-1;
}
```

Figure 8: Round-Robin partition strategy.

Table 3 shows an example of this allocation scheme for npt=20. We note a significant improvement in load balance. The difference in cell allocation and thus in load of all processors in the virtual machine has been reduced but not eliminated.

```
// MIRR partition strategy:  Given npt grid points and np processors, get rows
// assigned for processor Pi.  Rows are stored in the array rows.
getrows3(int Pi,int npt, int np,int *rows)
{
     int d,r,top,bottom, j=0, s=0, thisrow;
     if (np == 0 ) np = 1;
     d= (npt) / np ; r= (npt) % np ;
     top=Pi-np ; bottom=npt - 1 - Pi+ np;
     rows[j]=-1;
     while ( s < d ) {
       if ( s%2 == 0 ) { top += np; thisrow=top; }
       else { bottom -= np; thisrow=bottom; }
       rows[j]=thisrow;
       j++; s++;
     }
     if ( r !=0 && Pi< r ) rows[j++]= top+np;
     rows[j]=-1;
}
```

Figure 9: MIRR partition strategy

We further refine RR partition strategy to achieve further load balance with Mirror-Image Round-Robin partition strategy. MIRR allocates each processor P_i one row from the top and its corresponding complement from the bottom of the grid. The stub of figure 9 shows the algorithm used for this strategy. Figure 7(c) shows the cell allocation to each processor. MIRR has effectively achieved the load balance as shown in table 3

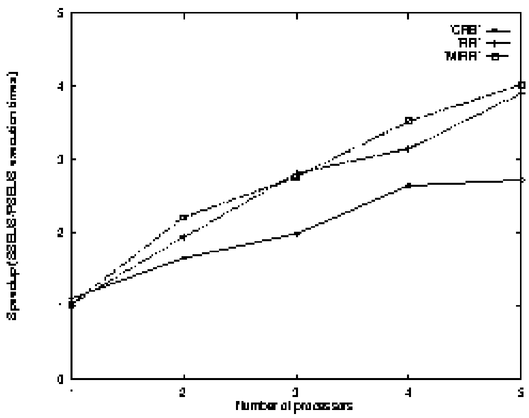


Figure 10: Relative speedup of PSEUS using three data partition strategies. The speedup is the ratio of PSEUS/SSEUS execution times for npt=10,000.

The speedup metric(Ratio of serial run times to the parallel run times) was used as a measure of performance.

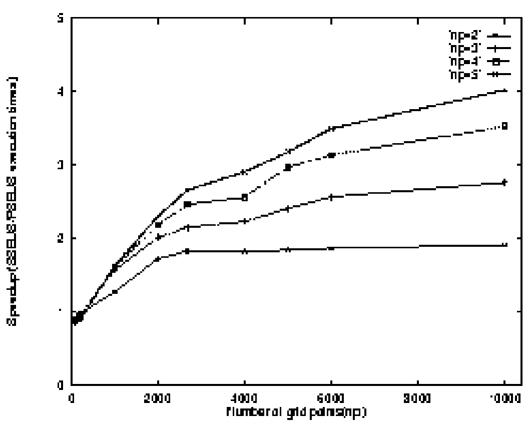


Figure 11: Speedup as a function of problem size and the number of processors for data MIRR partition strategy

Figure 10 shows the resulting increase in speed when PSEUS is distributed among 1,2,3,4 or 5 processors for a grid size of 10000×10000. RR and MIRR strategies result in a four fold speed increase, compared to the serial version. The increase in speed resulting from partition strategy RR and MIRR is significantly higher than that of CRB partition strategy. We also note that the increase in speed is a function of the problem size and the number of processors used(figure 11). Both RR and MIRR strategies however, require additional storage for row bounds since rows are not allocated in consecutive order. MIRR allocates rows in an unsorted order. This may impact locality when non-associative cache is used. The algorithm could be improved further by generating the rows in sorted order without additional overhead of a sort routine. CRB partition method has the disadvantage of load imbalance. However, this method may be useful in heterogeneous architectures. CRB or a variation of RR strategy (i.e., with different stride values) can be used in combination with a weight function to allocate a larger number of cells to faster slaves and a lower number of cells to slower hosts. It may also be used to change the number of cells allocated to a slave while it is executing by farming out portions of its computational domain when CPU load exceeds a threshold.

5 Conclusion

We have shown that some of the spatial interpolations and geographic informations systems analyses can be parallelized efficiently using distributed memory architectures. Various techniques from domain decomposition can be used to partition data among processors to achieve higher performance of applications. A new partition technique called Mirror-Image Round-Robin (MIRR) achieves load balance. Data partitioning algorithms should be considered carefully in order to achieve acceptable performance since there seems to be tradeoffs between the partition strategies. MIRR strategy may be suitable for homogeneous architectures, while CRB and RR strategies may be more suitable for heterogeneous architectures. A variation of the same techniques used in PSEUS could be applied to some of GRASS routines(e.g. to distribute data for CPU and I/O intensive operations). PSEUS is still a serial algorithm, we will be examining ways to change the computation algorithm itself to explore the parallelism. General purpose libraries for distributed spatial operations could ease the path towards a high performance distributed GIS architecture.

---

[1]National Remote Sensing Center, Office of Hydrology, NWS, 1735 Lake Drive West, Chanhassen, MN 55317 {elhaddi,mcmanamon,rhills}@nws.gov

[2]Department of Computer Science, University of Minnesota, 200 Union Street, Minneapolis, MN 55455 {elhaddi,shekhar}@cs.umn.edu

References

Anderson, 1973
Anderson, E. (1973). National Weather Service River Forecast System-Snow Accumulation and Ablation Model. Technical Report HYDRO-17, NOAA, NWS.

Carroll et al., 1995
Carroll, S., Day, G., Cressie, N., and Carroll, T. (1995). Spatial Modeling of Snow Water Equivalent Using Airborne and Ground-Based Snow Data. *Environmetrics*, 6:127-139.

Carroll, 1995
Carroll, T. (1995). GIS Used to Derive Operational Hydrologic Products From In Situ and Remotely Sensed Snow Data. In Carra, A. and Guzzetti, F., editors, *Geographical Information Systems in Assessing Natural Hazards*, pages 335-342. Kluer Academic Publishers, The Netherlands.

Carroll and Marshall, 1985
Carroll, T. and Marshall, R. (1985). Cost-Benefit Analysis of Airborne Gamma Radiation Snow Water Equivalent Measurements Made Before The February 1985 Fort Wayne Flood. In *Sixth Conference on Hydrometeorology, American Meteorological Society, Indianapolis, Indiana*.

Carroll and Vadnais, 1980
Carroll, T. and Vadnais, K. (1980). Operational Airborne Measurement of Snow Water Equivalent Using Terrestrial Gamma Radiation. In *48th Annual Western Snow Conference, Laramie, Wyoming*.

Carroll et al., 1985

Carroll, T., Vogel, R., and Gauthier, R. (1985). Operational Airborne Snow Water Equivalent and Soil Moisture Measuements Used in Water Supply and Snowmelt Flood Forecasting. In *Fifth Remote Sensing Symposium, Remote Sensing Applications For Water Resources Management, Ann Arbor, Michigan*.

Cressie, 1991
Cressie, N. (1991). *Statistics for Spatial Data* . John Wiley & Sons, Inc.

Day, 1990
Day, G. (1990). A Methodology for Updating Conceptual Snow Model With Snow Measurments. Technical Report 43, NOAA, NWS.

Fread et al., 1995
Fread, D., Shedd, R., Smith, G., Farnsworth, R., Hoffeditz, C., Wenzel, L., Wiele, S., Smith, J., and Day, G. (1995). Modernization in the National Weather Service River and Flood Program. *Weather and Forecasting*, 10(3):477-484.

Fritzsche, 1982
Fritzsche, A. (1982). The National Weather Service Gamma Snow System Physiscs and Calibration. Technical Report 43, EG& Energy Measurements Group, NWS-8201.

Geist et al., 1993
Geist, A., Beguelin, A., Dongarra, J., Jiang, W., Mancheck, R., and Sunderam, V. (1993). PVM 3 Users Guide & Reference Manual ORNL/TM-12 187.

Hartman et al., 1995
Hartman, R., Rost, A., and Anderson, D. (1995). Operational Processing of Multi-Source Snow Data. In *63rd Annual Western Snow Conference, Reno, Nevada*.

McManamon et al., 1995
McManamon, A., Hartman, R., and Hills, R. (1995). Implementation of The Snow Estimation and Updating System (SEUS) In the Clearwater River Basin, Idaho. In *63rd Western Snow Conference, Reno, Nevada*.

McManamon et al., 1993
McManamon, A., Szeliga, T., Hartman, R., Day, G., and Carroll, T. (1993). Gridded Snow Water Equivalent Estimation Using Ground Based and Airborne Snow Data. In *Proceedings of the 61st Western Snow Conference, Quebec City*, pages 75-81.

Peck et al., 1980
Peck, E., Carroll, T., and Vanemark, S. (1980). Operational Aerial Snow Surveying in the United States. *Hydrological Sciences-Bulletin-des Sciences Hydrologiques*, 25(3):51-62.

Shekhar and Liu, 1995
Shekhar, S. and Liu, D. R. (1995). Partitioning Similarity Graphs: A Framework for Declustring Problems. In *IEEE Intl. conf. Data Eng.*

Shekhar et al., 1995
Shekhar, S., Ravada, S., Turner, G., Chubb, D., and Kumar, V. (1995). Load Balancing in High Performance GIS: Partitioning Polygonal Maps. In *Lecture Notes in Computer Science*. Springer Verlag.

About this document ...

Mirror-Image Round Robin Spatial Data Partitioning :
a Case Study With Parallel SEUS

This document was generated using tex2html from cbl.leeds.ac.uk.
Please obtain a postscript version from ftp://ftp.nohrsc.nws.gov/pub/staff/elhaddi/mirr/mirr.ps
Please address all comments to elhaddi@nohrsc.nws.gov.

Brendan G. Mackey

THE ROLE OF GIS AND ENVIRONMENTAL MODELLING IN THE CONSERVATION OF BIODIVERSITY

GIS and environmental modelling provide new capabilities for analysing the space/time distribution of ecological phenomena. These predictive capacities are needed to supplement traditional descriptive inventory of Biodiversity. The primary environmental regimes provide a consistent framework for examining ecological process and pattern from global- to nano- scales. The multi-scaled nature of these driving processes requires an integration of climate modelling, terrain analysis, substrate and land cover data at specified scales of analysis. The key point is that ecological systems are hierarchically structured by the constraining influences of the primary environmental regimes. Analytical procedures based on these concepts provide the basis for the quantification and prediction of the potential niche of taxa. Also, Environmental Domain Analysis based on the same scale-dependent models of the primary environmental regimes, provides complimentary and essential context where biological data are limited.

INTRODUCTION

The conservation of Biodiversity remains one of the great challenges facing our generation and those to follow. The other great challenge is to sustain and nurture the globe's growing human population. The two are inexorably linked as economic development impinges upon the biota, while at the same time the biosphere provides essential resources for human well being. We should never forget that the loss of Biodiversity is a symptom of a failure to implement a process of sustainable development.

This paper discusses an operational framework for analysing terrestrial ecological systems from local to global scales and, in so doing, aims to define the role GIS and environmental modelling have to play in the conservation of Biodiversity. It is a well accepted principle that appropriately scaled information is required to underpin the complex decision-making needed to conserve Biodiversity. Information requirements are often interpreted as meaning simply *resource inventory*. However, knowing *what* is *where* is only the first step. Two other conditions must be met. First, an understanding is needed of the causal processes that lead to the extant pattern. From this can be built the capacity to predict what will happen given projected changes in land use and environmental conditions. The need for predictive in addition to descriptive natural resource inventory using computer-based methods was argued by Nix and Gillison (1985).

Meeting the above information requirements for Biodiversity has proven difficult for the following reasons:

(a) There is no single ecological unit of analysis, rather a variety of ecological phenomena are the foci of studies, including: populations; species; assemblages; communities; habitats; ecosystems; and biomes. It can be argued that *species* are the fundamental unit of conservation but the species concept is not applicable to all biota (for example, many species do not engage in sexual reproduction) and other levels of organisation both above and below a species are recognised as having significance to the conservation of Biodiversity (particularly populations, communities. and ecosystems). This has led to the lack of a standard methodology for inventorying the world's Biodiversity resources.

(b) The ecological world is very complex. The total number of species is large being estimated in the tens of millions (Wilson 1992). The total number of populations, communities and ecosystems is no doubt finite but possibly cannot be estimated. By comparison, ecological surveys are small in number, and the resultant quadrats or transects represent at best a scattered, irregular network of data points. While the ecological resources of these sites may be known in great detail, there remains the fundamental problem of spatially extending these data to cover the entire landscape or region. Neither do remotely sensed data provide a panacea for this problem. Whilst sampling the entire landscape, they provide no direct measure of ecological phenomena, rather the spectral/emittance data they record indicate surface conditions only and, in any case, must be interpreted by correlation with ground-based site data

(c) Information about the distribution and abundance of living organisms is in itself insufficient, as wild populations of plants, animals and micro-organisms depend upon components of the prevailing environmental conditions for their existence and continued adaption and evolution. Hence the habitat conditions for taxa are as important as data about their distribution and abundance

(d) Ecological systems are dynamic not static, that is, their number, type and condition of their components changes through time (species composition, biomass accumulation, population demographics etc.). Hence landscape must not only be surveyed for their Biodiversity resources, but be monitored through time to capture something of the system dynamics.

(e) Planning and management to conserve Biodiversity must deal with the distribution of these ecological resources through space and time, at a scale commensurate with that imposed on the landscape by human land use and resource extraction. Put simply, managers require lines on maps saying where things are, and what can or cannot be done with them.

While a definitive solution to the above impediments is perhaps a utopian goal, it is possible to implement an operational framework that would provide the basis for a consistent analytical approach applicable to all terrestrial ecological systems. This framework is built upon the integration of three fundamental concepts:

(a)an hierarchical approach to ecological phenomena

(b)the role of the primary environmental regimes in constraining biotic response

(c)the parametric approach to environmental survey and analysis.

The framework is considered to be operational because it utilises existing data and algorithms. All components have been applied in a variety of environmental settings (though they have yet to be all integrated into a single system). The following sections discuss how these concepts address the impediments noted above, and inso doing provide a context for defining the role to be played by GIS and environmental modelling in the analysis of ecological systems and hence the management and conservation of Biodiversity.

AN HIERARCHICAL APPROACH TO ECOLOGICAL PHENOMENA

Allen and Hoekstra (1992) argued that the multitude of ecological units noted above in point (a) are better thought of as criteria that can be applied to distinguish different forms of ecological phenomena. It was noted that they are largely scale independent so that all criteria can be applied at any scale. Fig. 1 shows an interpretation of the conventional nested ecological hierarchy as represented by organism-population-species-etc. Allan and Hoekstra argued against such a schema by rejecting the notion that ecological phenomena represent a nested hierarchy, ie. that ecosystems contain communities which contain populations etc. For example, as illustrated in Figure 1, ecosystem processes can be defined based on the community of organisms living in a cow's stomach. Similarly, the life history of some whale species carry them around the globe. The key is to recognise that (a) *species* are not the only valid ecological units of study, and (b) the scale of observation is a major factor determining how the unit of study is defined. The confusion arises, at least partly, because living organisms interact at different scales with their environment. Anyone who has worked with wildlife is familiar with this notion via the concept of home range. For example, It is intuitively obvious that an elephant interacts with the environment at a larger scale than does a field mouse (also note that variations in the scale of environmental interactions also occur with plants). It follows that more than one scale of observation is required to fully capture the ecological resources of a landscape or region. Ecological surveys tend to focus on only one of the ecological criteria with the selected criterion being applied at a particular scale. Therefore all field data are limited as multiple criteria need to be applied at a range of scales to fully account for the distribution of Biodiversity.

The cross-scaled nature of ecological phenomena is further compounded by the multi-scaled nature of the environmental processes that constitute an organism's habitat. We are dealing with complex systems. Traditional scientific methods of inquiry simplify matters by holding most processes constant, and changing one variable at a time to assess system response. Not only is this intellectually unsatisfying but the complexity bought about by scale-related issues lies at the heart of ecological phenomena. This complexity is something that must be explicitly dealt with if the phenomena are to be adequately understood.

Hierarchy Theory (HT) was formalised by Koestler (1967) as an alternative to the application of classical reductionist experimental science in the study of human behaviour. Koestler argued that the results of experiments on rat behaviour could not be extrapolated to humans, because of the latter's greater inherent complexity. HT was proposed as a conceptual framework for dealing with complexity as a phenomena *per se*. HT was interpreted and articulated in an ecological context by O'Neil et al. (1986) and Allan and Starr (1988). Certain concepts of HT are relevant to the present discussion:

(a) Systems are hierarchically structured if processes operating at a larger scale constrain processes operating at a smaller scale; the key word is *constrain* (NB: large scale is defined as low frequency [slow turnover] and geographically extensive, while small scale is defined as high frequency [fast turnover] and geographically restricted)

(b) Each level within a hierarchy is a component of a higher (larger scaled) level, while simultaneously sets the context for the level beneath it (that is, the smaller scaled phenomena). Koestler defined the term *holon* to reinforce this two-faced characteristic of hierarchical systems.

(c) Holons generate emergent properties which are not apparent from an analysis of the individual components, that is, the whole is greater than the sum of the parts.

Given this, the challenge is to develop a framework that will enable the multi-scaled nature of biota and their environments to be dealt with at the surface of the landscape where they intersect with human land use and resource extraction. The key questions then become: can we identify a fundamental set of generic processes that generate the hierarchical structures found in ecological systems - processes that provide, at a range of scales, the fundamental set of constraints within which biological systems operate; and are these amenable to inventory and analysis using GIS and environmental modelling ?

THE PRIMARY ENVIRONMENTAL REGIMES

Defining the effective environment experienced by living organisms is a long standing objective of ecological studies (eg. Waring and Major 1964). As noted above, lack of a common or standard methodology can be attributed to a number of factors including that studies focus on different ecological criteria at different scales. However it is generally accepted that all photosynthesising plants are influenced by four primary environmental regimes, viz. radiation, thermal, moisture and mineral nutrient. Each regime comprises processes that determine the distribution and availability of light, heat, water and mineral nutrients. These resources exert a fundamental control on plant physiology and related functions. The primary environmental regimes also directly and indirectly affect animal response. For example, animal species are adapted to optimum thermal conditions in an analogous manner to plants. Indirectly, the regimes influence animals via their control on plant growth; a source of shelter and nutrition for many species.

The relations between the primary regimes and living organisms have been the subject of many ecological studies. However the significance of the fact that each regime represents a suite of processes is frequently overlooked. The implication is that there is no single scale at which the primary environmental regimes can be defined. Rather, the scale of analysis determines which processes are considered in defining the regimes. Figure 2 shows the five major scales at which the regimes can be modelled. This schema is adapted from Linacre (1992) (who was arguing for consistency amongst climatologists in dealing with scale). The figure also shows examples of the processes and factors that operate at each scale, and that can be used to define the primary environmental regimes.

Traditionally, biogeographers have operated at global- to meso-scales. Forexample, the work of Holdridge (1967) and Box (1981) made fundamental contributions in correlating biomes/vegetation structure with globally scaled climatic parameters related to temperature and moisture. Most ecological work undertaken this century however has been focussed at the micro- and nano-scales. If we temporarily limit our discussion to forested ecosystems, we can identify two processes that dominate at these small scale: at the micro scale, the effect of canopy cover in modifying light, temperature and moisture conditions for understorey organisms; at the nano scale, the role that soil micro-organisms and the vegetation play in cycling and conserving nutrients, thereby acting as a break on the leakage of nutrients from the system by leaching and erosion.

Both climatic and geological processes dominate at the meso-scale. The spatial and seasonal distribution of the prevailing weather systems is critical. To this we can add the interactions of these systems with elevation, and the resultant lapse rate effects on temperature and precipitation (ie. temperature is cooler at high elevations, and precipitation increases, up to a point and all other factors being equal, with elevation). These processes are important as long term, mean monthly climate provides the basic inputs of precipitation, insolation and air temperatures into the primary environmental regimes. The other meso-scale process relates to the role played by the geological substrate as a primary source of minerals, and the control this exerts on soil chemistry and hence the mineral nutrient regime. Though it is important to remember that the other primary source of minerals is the atmosphere as nutrients can enter the landscape via rainfall attached to particles. Golley (1983) show for tropical rainforest that given sufficient time, and efficient nutrient cycling and conservation mechanisms, vegetation systems can obtain all the nutrients they need via precipitation. Hence rainforest is commonly found on nutrient poor geological substrates (also see Walker et al. 1981). This can be interpreted as an example of micro-scaled processes exerting control over the constraints set by larger scaled meso-processes.

The topo-scale as defined here refers to two main processes, First, the role played by the topography in redistributing water in the landscape. Second, the effect of slope, aspect and horizon shading in modifying the amount of solar radiation received at the surface. The former is a major factor determining the moisture regime. The latter has both direct and indirect effects. Certain plant taxa have a competitive edge in terms of being able to germinate and grow under low light. For example, in Australian temperate forests, the suite of tree taxa collectively called *rainforest*, out-compete members of the sclerophyll *Eucalypt* genus in shaded locations (eg. Barret and Ash 1992). Reduced radiation also results in decreased evaporation rates (all other factors being equal) and hence higher soil moisture levels, in addition to lower air, surface and soil temperatures.

The PER schema outlined in Figure 2 also provides a framework for dealing with the impact of disturbances in the landscape. Wildfire in forest ecosystems is generally a large scale phenomena (big extent, relatively infrequent) but its impact on ecological systems is at a smaller scale. Fire does not change the PERs at the meso- and topo-scales, however it does impact on the micro-scale processes associated with the vegetation canopy. Similarly, forestry operations modify thePERs at both the micro- and nano-scales through their effects on vegetation cover and the soil profile.

THE ROLE OF GIS AND ENVIRONMENTAL MODELLING

Far less ecological research has been undertaken at the meso- and topo-scales compared to the global-, micro- and nano-scales. In fact it is only since the relatively recent development of certain computer-based analytical methods that it has been possible to effectively deal with the primary environmental regimes at these scales. GIS-related techniques have been critical as the problem has been to deal with these processes in a spatially distributed manner.

For example, the climatic interpolation and gridding work of M.F. Hutchinson, H.A. Nix, J.P. McMahon and colleagues (Hutchinson et al. 1984, Nix 1986) has provided a hitherto unavailable capacity to generate spatially reliable estimates of the meso-scaled climatic inputs to the primary environmental regimes. These have been applied at the meso-scale using a variety of ecological criteria. The advantages of this method stem partly from the explicit incorporation of additional variables to that of the position of the climate stations as independent variables in the surface fitting procedure. This means that elevation lapse rate effects can be effectively dealt with, resulting in standard errors for temperature predictions of around 0.5 degrees Celsius, and for precipitation of 10-15%.

Similar GIS-related developments have transformed our ability to spatially model processes at the topo-scale. The advent of digital elevation models and associated terrain analysis algorithms means that the topo-scaled effects on the primary environmental regimes can be now factored in. The TAPES suite of software by Ian Moore and colleagues (Gallant and Wilson *in press* 1996, Wilson and Gallant *in press* 1996, Wilson and Gallant *in review* 1996) provide the capacity to account for both the catchment hydrology and radiation shading processes that dominate in erosional, humid forested landscapes (see Moore et al. 1993, Mackey 1994, Mackey et al. 1995). The limiting factor becomes the availability of suitably scaled indices of the PERs. While critical processes at the meso- and topo-scales can now be captured, the reliability of spatial data decreases at the micro- and nano-scales.

The main sources of spatial data about the substrate remains thematic maps of bedrock and surficial geology (though developments in the application of remotely sensed radiometric data are promising). However these are generally derived for prospecting and engineering purposes, and the mapped units may have little direct interpretation in terms of the mineral nutrient regime experienced by plants. One approach is to reclassify the mapped data to reflect the relative potential of each class as a source of potential nutrients assuming soil has developed *in situ* on this parent material. This can be done qualitatively with the aid of an experienced geologist, or at least semi-quantitatively if adequate rock chemistry data are available. In this way, the meso-scaled influence of parent material lithology can be coupled to similarly scaled climate models to define spatial estimates of the PERs inclusive of the mineral nutrient regime. This approach has been used in a variety of studies in Australian forested landscapes including Webb (1968), Nix (1993), Mackey et al. (1989), JSC (1990), Mackey (1993)).

The increased capacity provided by GIS and spatially distributed environmental modelling enables two fundamentally new analytical capabilities.

Spatial Modelling of Physiological Niche

The first stems from the ability to estimate *ex situ* at survey sites indices of the primary environmental regimes. For example, the climate modelling work of Hutchinson and colleagues noted above enables estimates of long term mean monthly climate to be generated at remote field sites. Hence the potential climatic domain occupied by taxa can be empirically determined. By coupling a digital elevation model to these climate surfaces, gridded estimates of the climatic parameters can generated. Matching the gridded data to the taxa's climatic profile enables the potential domain to be spatially predicted. (see Nix 1986, Booth 1985, Busby 1986). Meso-scaled modelling and/or spatial prediction of the potential niche of plant and animal species can now be undertaken that incorporate topographic and substrate factors in addition to long term mean monthly climate, utilising a variety of statistical modelling approaches to correlate biotic distributions or abundances with these indices of primary environmental regimes.

Figure 3 shows how analysis of a taxa's physiological niche can assist in understanding its ecological niche. By modelling the response of taxa to the primary environmental regimes, a spatial prediction of its physiological niche can be generated. This provides an environmental and geographic context for exploring the other processes that exert control on the distribution and performance of living organisms: evolutionary history; disturbance; biological interactions; present and past land use. For example, Mackey et al. (1989) compared the potential distribution of rainforest structural types to extant rainforest cover (derived from remotely sensed data) in order to assess the conservation significance of the remnant vegetation patches in the landscape.

However, a major problem facing the conservation of Biodiversity is the paucity of biological data. We have previously noted that wild populations of living organisms are dependent upon varying suites of environmental factors that relate to both the physical environment and other biota. Quantifying the ecological niche for a taxa requires long term (~10 years) of field study, complimented by computer-based modelling and statistical analysis (eg. Lindenmayer et al. 1993). Given the vast number of species and genetically differentiated populations on the planet, the most we can expect is to have modelled a rough approximation of the major components of habitat for a few of the major vascular plant and vertebrate taxa. Most species will become extinct before their habitats can be defined, let alone before their intra-species, population-based diversity is examined.

Environmental Domain Analysis

One solution lies in identifying landscapes that are likely to be significantly different in term of their Biodiversity, and to ensure that representative samples of these landscapes are protected either within conservation reserves, or through control of land management practices. Classifying and delineating landscapes based on the ecological resources requires data in a spatially distributed form. For example, vegetation data can be used as a surrogate for unknown taxas' distributions andhabitats. This results in what Clifford and Stephenson (1975) called a *biocoenotic* classification, that is, where discontinuities in the biota are used to divide the landscape. The alternative is a *biotopic* classification where the spatial distribution of biota are defined based on discontinuities in the abiotic environment. Hence *ecoregionalisations* can be mapped using either surrogate biological data or based on some type of physiographically-based criteria such as catchment boundaries.

An alternative method was proposed by Mackey et al. (1988) which has subsequently become known as Environmental Domain Analysis (EDA). The method has been widely applied and significant developments made by various authors (see Mackey et al. 1989, JSC 1990, Lewis et al. 1991, Belbin 1993, Mackey et al. in press). EDA drew upon the environmental modelling methods developed by H.A. Nix and colleagues noted above. The basic methodology is outlined in Figure 4. The first step involves generation of the primary climatic, terrain and substrate data. From these are generated gridded values for selected indices of the PERs. Agglomerative cluster analysis is then applied to the gridded data that compares all grid points (cells) to each other based on their PER index values using an appropriate metric. In this way the relative similarity of each grid point to every other grid point is quantified. Grid points are then clustered into groups that share relatively similar values for the PER indices. Any number of groups can be identified from *1* to *n* (the total number of grid points)and their relative inter-group similarity represented as a hierarchy. As the location of each grid cell is implicitly recorded in a raster-based GIS, a map of the classification can be produced based on any desired level of similarity.

EDA fundamentally differs from the traditional notion of regionalisation. It is based upon the explicit ecological framework provided by the primary environmental regimes, and utilises computer-based environmental modelling and analysis techniques. Regionalisation implies the mapping of large scale, homogeneous areas. EDA will identify clusters of grid cells that have similar environments irrespective of their lack of adjacency. Hence both outliers and inliers to the broader environmental pattern can be identified.

It follows that there are two dimensions to EDA. The first refers to the hierarchy of scales at which the processes controlling the PERs can be modelled (Figure 2). The second encompasses the hierarchy of classification groups that can be represented at a given scale. Taken together, they provide a consistent framework for examining ecological pattern and process from global to local scales. To date, applications of EDA have been at the meso-scale, however on-going research at the Australian National University (R.G. Lesslie) and the Canadian Forest Service (D.W. Mckenney and colleagues) is applying the theory at the topo-scale.

CONCLUSION

GIS and environmental modelling provide the means to directly model the potential habitat of specific taxa, so long as adequate empirical data about their distribution and productivity are available to calibrate their response to the PERs. Hence great advances are now possible in the spatial analysis of a taxa's physiological niche, and the incorporation of other data needed to estimate its ecological niche.However lack of appropriate biological data in light of the pace at which resource utilisation is occurring, demands complimentary approaches such as that provided by EDA.

As noted above, taxa interact with the environment at different scales; hence populations, species, communities and ecosystems can all occur at smaller or larger spatial scales than each other. This observation seriously undermines the value of ecological regionalisations. It is possible to successively subdivide spatial units to generate, for example, a nested spatial hierarchy. Various countries have developed such classifications and are using them to assist in evaluating the conservation status of Biodiversity (eg. Thackway and Cresswell 1995). The problem is that there is not a similarly nested set of ecological phenomena whose distributions coincide with these boundaries. Regions so defined will certainly contain some ecological phenomena, and may in fact be dominated by a particular phenomenon. Also in certain cases they will contain the entire distribution of a particular phenomenon. However the only spatial boundary that absolutely contains living organisms is the atmosphere surrounding planet Earth. All other boundaries delineated on the earth's surface will to some degree be *permeable* to the flow and distribution of some or other ecological phenomena. Consequently there is no such thing as the optimum ecological regionalisation. Rather the problem is how to deal with the bio-environmental continuum that exists within the spatially bounded planetary ecosystem.

EDA provides a method for dealing with the planetary continuum in a way that fundamentally changes how we think about ecological regionalisation. It is possible to undertake EDA at any of the scales (from global to nano) so long as spatially distributed estimates of the PERs are available. The problem is that as we progress down the hierarchy (from large to small) it becomes increasingly difficult to generate spatial data to model the processes operating at the smaller scales. Consequently, applications of EDA have focussed on the meso-scaled processes, thereby taking advantage of computer-based GIS and environmental modelling techniques for generating gridded estimates of long term mean monthly climate, and integrating these with digitised mapped geological data. But given the advent of digital terrain analysis based on 20-30 resolution elevation models (and their use in simulating catchment hydrology and generating gridded estimates of radiation), the possibility is now open to undertake EDA at the topo-scale. Remotely sensed satellite imagery clearly has a vital role to play in the provision of vegetation cover data as input to analysis of the PERs at the micro scale.

The multi-scaled nature of living organisms and primary environmental regimes means that the ecological volume of a landscape is far greater than its surface area. Many of the dimensions that define this ecological volume may be hidden from human perception when viewed by the unaided primary senses. GIS and environmental modelling open new windows of observation that extend our perception into these hidden dimensions. The end result is a vastly improved capacity to inventory the distribution of the world's Biodiversity through space and time. Analysing ecological relations within the context of the primary environmental regimes provides a consistent framework for understanding the causal processesunderpinning extant patterns of ecological phenomena. This in turn provides the basis for prediction of the likely impact of environmental change on the conservation of Biodiversity.

ACKNOWLEDGMENTS

I am grateful to Val Lyon, Department of Geography ANU, for creative input to and preparation of Figures 1-4.

REFERENCES

Allan T.F.H., and T.B. Starr (1992). *Toward a unified ecology*. Columbia University Press, NY, 384 pp.

Allan T.F.H., and Starr T.B. (1988). *Hierarchy: perspectives for ecological complexity*. The University of Chicago

Press, Chicago.

Barret D.J., and Ash J. (1992). Growth and cabon partitioning in rainforest and eucalypt forest species of south coastal New South Wales, Australia. *Australian Journal of Botany* 40:13-25.

Belbin L. (1993). Environmental representativeness, regional partitioning and reserve selection. *Biological conservation* 66:223-230.

Booth T.H. (1985). A new method for assisting species selection. *Commonwealth Forestry Review* 64:241-250.

Box E.O. (1981). *Macroclimate and plant form*. Tasks for vegetation science 1. Dr. W. Junk Publishers, The Hague.

Busby J.R. (1986). A biogeoclimatic analysis of Nothofagus cunninghamii (HOOK) Oerst in eastern Australia 1. *Australian Journal of Ecology* 11:1-7.

Clifford H.C., and Stephenson W.C. (1975). *An introduction to numerical classification. Academic Press*, New York.

Gallant J.C., and Wilson J.P. (in press, 1996). A grid-based terrain analysis program for the environmental sciences. *Computer and Geosciences* 22(4)

Golley F.B. (1983). Nutrient cycling and nutrient conservation. IN. F.B.Golley (editor). *Tropical rain forest ecosystems: structure and function*. Elsevier Scientific Publishing Company, New York, pp.137-156.

Holdridge L.R. (1967). *Life zone ecology*. Tropical Sciences Centre, San Jose.

Hutchinson M.F., Booth T.H., McMahon J.P., and Nix H.A. (1984). Estimating monthly mean values of daily solar radiation for Australia. *Solar Energy* 32:277-290.

JCS (1990). *Joint Scientific Committee report on biological conservation in the south east forests to the Hon. Alan Grifiths, MP, Minister for Resources, Commonwealth of Australia, and the Hon. Ian Causley, MLA, Minister for Naturalresources, State of New South Wale*s. Australian Government Publishing Service, Canberra.

Koestler A. (1967). *The ghost in the machine*. Macmillan, New York.

Lewis A., Stein J.L., Nix H.A., Mackey B.G., and Bower J.K. (1991). An assessment of regional conservation adequacy, Tasmania. *Forest and Timber Inquiry Consultancy Series No. FTC91/17*. Resource Assessment Commission, Canberra.

Linacre E. (1992). *Climate data and resources: a reference and guide*. Routledge, London, 366 pages.

Lindenmayer D.B., Lacy R.C., Thomas V.C., and Clark T.W. (1993). Predictions of the impacts of changes in population size and environmental variability on Leadbeater's Possum, Gymnobelideus leadbeateri McCoy (Marsupialia: Petauridae) using Population Viability Analysis: an application of the computer program VORTEX. *Wildlife Research* (1993) 20:67-86.

Mackey B.G., McKenney D.W., Yin-Qian Y, McMahon J.P. and Hutchinson M.F. (in press). A climatic analysis of Hills' site regions for the Province of Ontario. *Canadian Journal of Forest Science*.

Mackey B.G., Sims R.A., Baldwin K.A., and Moore I.D. (1995). Spatial analysis of boreal forest ecosystems: results from the RInker Lake case study. *Proceedings of the second international workshop/conference on GIS and environmental modelling, Breckenridge, 1993*. National Centre for Geographic Information and Analysis, University of California, Santa Barbara.

Mackey B.G. (1994). Predicting the potential distribution of rainforest structural types. *Journal of Vegetation Science* 5:43-54.

Mackey B.G. (1993). A spatial analysis of the environmental relations of rainforest structural types. *Journal of Biogeography* 20:303-336.

Mackey B.G., Nix H.A., McMahon M.F., and Fleming P.M. (1988). Assessing the representativeness of places for conservation reservation and heritage lisitng. *Environmental Management* 12:501-514.

Mackey B.G., Nix H.A., Stein J.A., and Bullen F.T. (1989). Assessing the representativeness of the Wet Tropics of Queensland, World Heritage Property. *Biological Conservation* 50:279-299

Moore I.A., Norton T.W., and Williams J.E. (1993). Modelling environmental heterogeneity in forested landscapes. *Journal of Hydrology* 150:717-747.

Nix H.A. (1993). An environmental framework for Australian rainforests. IN. G.L. Werren and A. K.Kershaw (editors). *The rainforest legacy*. Australian Heritage Commission. Special Publication 7(2). Australian Government Printing Service, Canberra.

Nix H.A. (1986). A biogeographic analysis of the Australian elapid snakes. IN. R Longmore (editor). *Atlas of elapid snakes*. Australian flora and fauna series, number 7, pp.4-15. Australian Government Publishing Service, Canberra.

Nix H.A., and Gillison A.N. (1985). Towards an operational framework for habitat and wildlife management. IN. J. Kikkawa (editors). *Wildlife management in the forests and forestry-controlled lands in the tropics and the southern hemisphere*. Proceedings of a workshop held at the University of Queensland, July 1984. University of Queensland, Brisbane, pp.39-55.

O'Neil R.V., DeAngelis D.L., Waide J.B. and Allen T.F.H. (1986). *A hierarchical concept of ecosystems*. Princeton University Press, New Jersey.

Thackway R., and Cresswell I.D. (1995). *An interim biogeographic regionalisation for Australia*. Australian Nature Conservation Agency, Canberra.

Walker J., Thompson C.H., Fergus I.F. and Tunstall B.R.(1981). Plant succession and soil development in coastal sand dunes of subtropical eastern Australia. In. R.C. Mcdonald, H.H. Shugart, D.B. Botkin (editors). *Forest succession: concepts and applications*. Springer-Verlag, New York, pp.107-131.

Waring R.H., and Major J. (1964). Some vegetation of the California coastal redwood region in relation to gradients of moisture, nutrients, light and temperature. *Ecological Monographs* 34(2):167-215.

Webb L.J. (1968). Environmental relations of the structural types of Australian rainforest vegetation. *Ecology* 49(2):296-311.

Wilson E.O. (1992). *The diversity of life*. W.W. Norton Company and Inc., New York, 424 pages.

Wilson J.P., and Gallant J.C. (in press, 1996). EROS: a grid-based program for estimating spatially distributed erosion indices. *Computers and Geosciences* 22(4).

Wilson J.P., and Gallant J.C. (in review). SRAD: a program for estimating radiation and temperature in complex terrain. *Computers and Geosciences*.

Figure captions

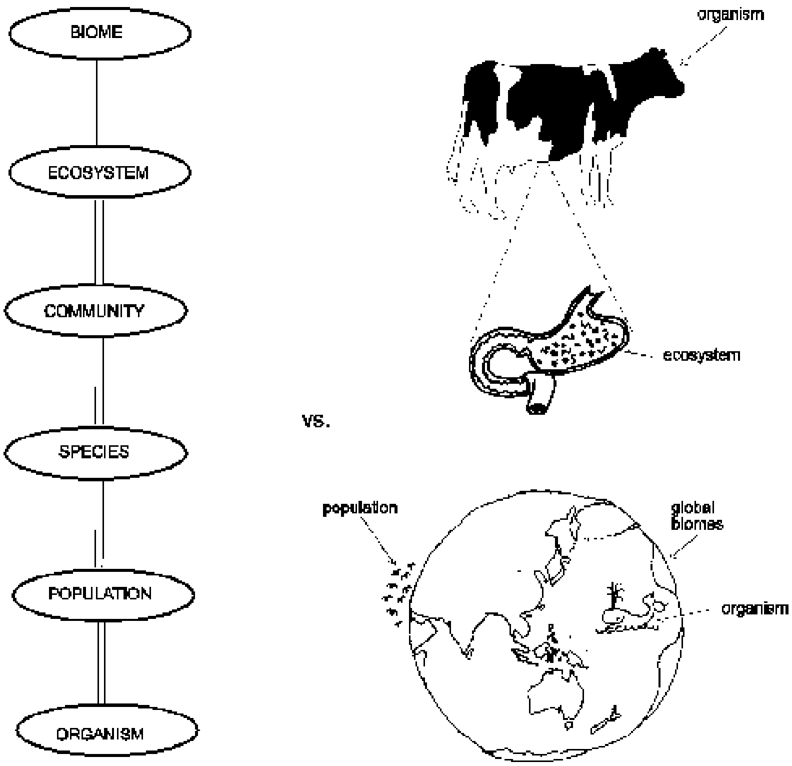


Figure 1. Conventional notions of a nested ecological hierarchy do not match the cross-scale reality of ecological phenomena (adapted from Allan and Hoekstra 1992).

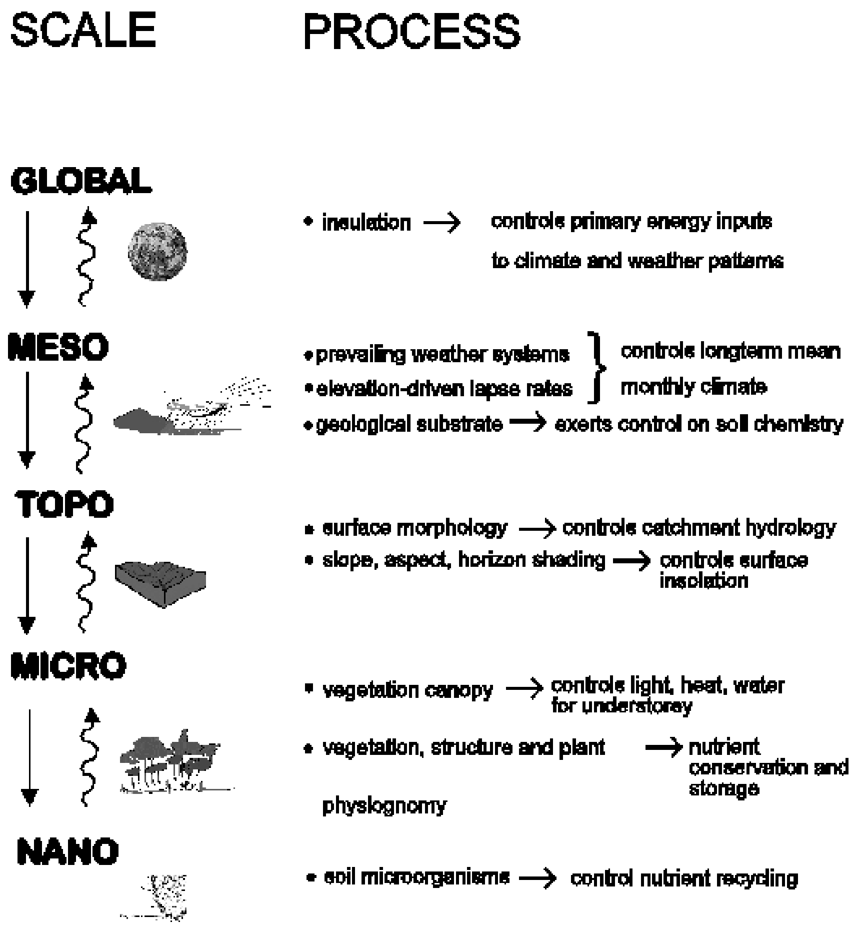


Figure 2. Scales at which various biophysical processes dominate calculation of the primary environmental regimes. While there is some feedback up the hierarchy,each level is dominated by constraints imposed at the larger scale. Note that at micro-scales, biologically-mediated processed dominate in defining the effective environment for a given organism. This schema assumes the focus of analysis is a forested site in a humid, erosional landscape (modified from a climatic schema by Linacre 1992). Additional important processes occur at all scales.

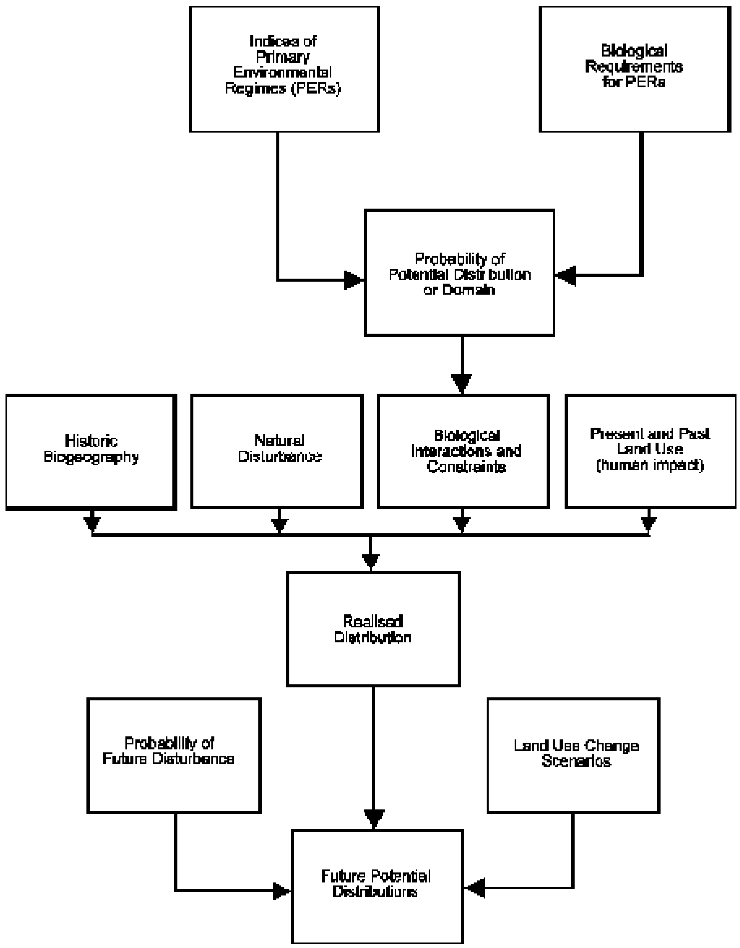
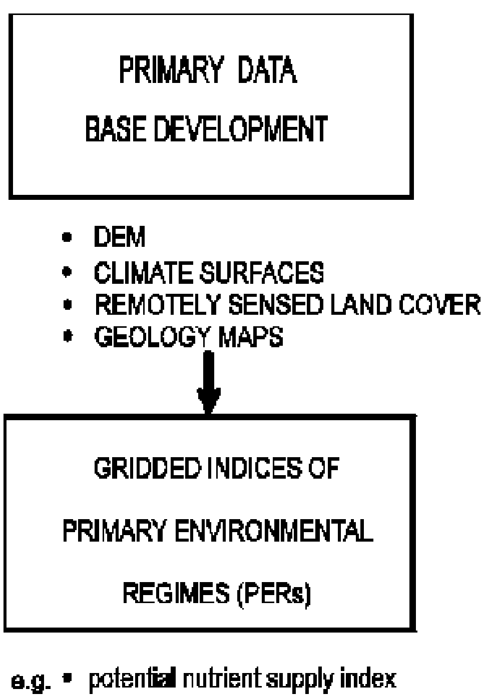


Figure 3. Relationship between processes that determine a taxa's potential (physiological) and actual (ecological) niche. These relations provide the basis for predicting future ecological response.



- mean annual temperature
- total annual precipitation
- precipitation seasonality

RELATIONAL FILE OF PERs

grid cell position	potential nutrient supply index	mean annual temperature	total annual precipitation	precipitation seasonality
$x_i y_i$	6	12	1000	.5
⋮	⋮	⋮	⋮	⋮
$x_N y_N$	7	8	2000	.1

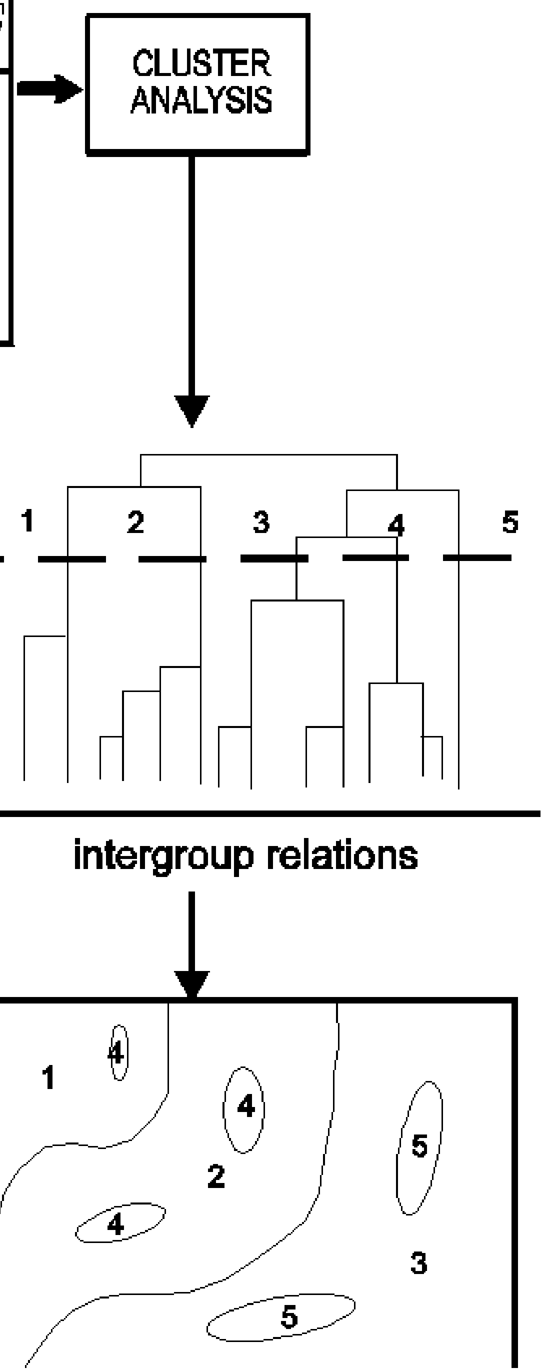


Figure 4. Main analytical steps in Environmental Domain Analysis. Note that groups can be defined and mapped at any level in the hierarchy.

Author information: Dr. Brendan G. Mackey, Department of Geography, School of Resource Management and Environmental Science, Australian National University, Canberra ACT 0200, Australia. (email: brendan.mackey@anu.edu.au; fax: 61-6-249 3770)

Richard Church, David Stoms, Frank Davis, and B.J. Okin

Planning Management Activities to Protect Biodiversity with a GIS and an Integrated Optimization Model

Abstract

We present the details of a general spatial model that was developed for the selection of biodiversity management areas in the Sierra-Nevada Region. This model is loosely integrated with a GIS system. The basic modeling approach begins by first identifying those plant communities that are vulnerable due to land use activities in current management plans. The level of vulnerability is assessed for each element of interest on a spatial basis using ARC/INFO. The planning problem involves selecting an efficient set of watersheds for biodiversity management through specially developed heuristics and the Optimization Subroutine Library of IBM. Results of this approach are given for the northern region of the Sierra Nevada of California. The BMAS model represents a significant advance in GIS-based conservation planning, both in sophistication of the algorithms used and in the integration of cultural and land use data with biological data.

Introduction

Protecting biodiversity has been the subject of a number of research projects and will continue to be of interest worldwide. Such interest is predicated on the loss of species due to land use changes and habitat losses as well as estimates for global climate change and increased variation within local climates. Researchers have discussed a variety of methods to protect habitats and biodiversity within a region; such methods include developing reserves and corridors, changing land use patterns, managing the landscape within limits to protect biodiversity, and managing captive breeding and release programs. Within a region, practically all of these techniques might be used to protect the natural environment and its diversity. To address regional biodiversity and plan for its protection, it is necessary to maintain up-to-date and accurate data on a wide variety of data coverages. Because of the detail and amount of such data as well as the importance of the spatial relationships inherent in location and proximity, it is natural to take advantage of GIS systems. We view the use of GIS as not only valuable, but for all intents and purposes, indispensable.

Much of the Sierran Region has been subjected to primary economic activities involving timber harvesting, grazing, recreation, mining, and water resource development. The impacts of such activities have been considerable. For example, aquatic biodiversity has been severely impacted by water impoundments and diversions, logging and grazing of riparian habitats, and the introduction of non-native species (Moyle and Randall 1996). In the Sierran foothills, long term (>150 years) livestock grazing has profoundly altered herb layer composition, aided the invasion of exotic plant species, degraded or eliminated riparian vegetation, and reduced soil fertility.. Ongoing rapid urbanization is spawning an extensive and possibly unmanageable urban-wildland interface. Even though middle and upper montane westside forests are relatively intact there is very little structurally complex forest left (11% of total). For the eastside of the Sierras, late seral conditions remain on only 7% of this forest type. Montane and subalpine meadows have been subject to long term sheep and cattle grazing. The legacy of grazing has altered soils, vegetation composition and hydrology in many meadow ecosystems.

We define Biodiversity Management Areas as specially designated public or private lands with an active ecosystem management plan in operation whose primary purpose is to contibute to regional maintenance of native genetic, species and community levels of biodiversity, and the processes that maintain that biodiversity. The primary management goal in each BMA is to sustain native biodiversity. The purpose of this paper is to describe the model used to select BMAs.

The Overall BMA strategy

Biological conservation strategies have traditionally centered on biological reserves, where a reserve is "an area with an active management plan in operation that is maintained in its natural state and within which natural disturbance events are either allowed to proceed without interference or are mimicked through management. The viability of a reserve system is measured on the size, shape, and connectedness of these remnant habitat areas.

In much of the Sierra Nevada large portions of public and private land are managed for renewable natural resources such as livestock forage, timber, and recreation. The prevailing land cover types of the Sierra Nevada are managed forest, rangeland and alpine ecosystems that sustain many if not most elements of native biodiversity while also supporting natural resource-based economies. Thus, a Sierran BMA system would not function as an archipelago of biological reserves. However, a BMA system could provide "core habitat areas" of higher habitat quality for many species and/or sanctuaries for species and habitat types that are negatively impacted by human activities.

The BMA strategy is not intended to provide a comprehensive reserve system for the Sierra Nevada that will ensure against the extinction of species or ecosystems (Davis et al. 1996). Developing and evaluating such a system is beyond the scope of available data and resources, and would require us to make many assumptions about ecosystem processes and management of non-BMA lands that would be tenuous at best. Instead, we view instituting a system of biodiversity management areas that represent all major ecosystems as just one component of an overall biodiversity conservation strategy for the Sierra Nevada.

A number of questions need to be answered in order to develop the best startegy for biodiversity management in the Sierras. Such questions include:

1. *What is the minimal area required to represent all Sierran plant community types in*

BMAs? How does an "optimal" BMA system compare to the existing set of parks, wilderness areas and reserves in the region?

2. *Can a representative BMA system be established on public lands only? If not, what area of private lands is required? How does the area requirement change if lands that are currently administratively withdrawn from grazing and timber harvest are classified as BMA lands?*
3. *How sensitive is the siting of BMAs to the way in which biodiversity is measured? Specifically, how do solutions to represent plant community types compare to solutions based on representing vertebrate species?*

Through the use of the model presented in the next section, we have addressed these issues and others in the SNEP study.

A Model for Selecting Biodiversity Management Areas

Assume that we have forecast both land use and activities for a region. This forecast would include land use changes by small watersheds or planning units. Given land use and the presence and extent of communities, habitats, and species, we can determine the extent to which various species, communities, endemic plants, old growth stands, etc. (called elements) are at risk now or in the future. We will call such elements at risk as being vulnerable. In order to reduce vulnerability, it is necessary to select BMAs in such a manner that sufficient distribution of that element is included among the selected BMAs. Selecting enough area for a vulnerable element, may mean that several watersheds or planning units need to be selected for BMAs. Some areas may be more suitable for selection than others. For example, if an area already has a dense human population or extensive road network and depending on the vulnerable element, the area may not be very suitable for targeting it for biodiversity management activities. Consequently, we would want to select those areas that are most compatible with our objective of improving the future of a given element and most compatible with existing land use.

It is necessary to select enough area for biodiversity management options that we keep elements from being considered vulnerable. Since we might consider hundreds of elements to be vulnerable and we can select from among hundreds of planning units for targeted action, the problem is relatively complex. We can represent this decision problem as an integer-linear programming model where the objective is to optimize the selection of suitable areas for biodiversity management such that enough area is selected for each element to keep it from being considered vulnerable. In order to formulate this model consider the following notation:

Notation:

In order to formulate this model consider the following notation:

j, J	index and set of planning units (e.g. small watersheds)
k, K	index and set of vulnerable biodiversity elements. These elements can be communities, species, endemic plants, old-growth stands, etc. They have been identified as being vulnerable without intervention of some sort, which includes changing management practices or land use zoning.
a_j	the area of planning unit j
Hd_j	human density measurement for planning unit I_j. The higher the value the less likely the area can be an effective biodiversity management area.
r_j	the percent of the area in planning unit j that is impacted by roads. The higher the value, the more development/access and the lower the suitability for using as a biodiversity management area.
Pla_j	the percent of the area of unit j that is held in private ownership. The higher the percentage, the more costly it would be to manage for biodiversity.
PPI_j	the density of public-private land interface. This is captured as the length of public-private land boundary in planning unit j divided by the area of the planning unit. The greater the length interface for a given area, the more fragmented the public land and the less suitable for use as a biodiversity management area.
a_{jk}	area in planning unit j which contains element k and is potentially impacted by current or planned activity in j
Min_k	minimum area containing element k that needs to be brought under biodiversity management in order to remove element k from the list of vulnerable elements
W_l	weight attached to term l in the objective function
Xj	$\begin{cases} 1 & \text{if site } j \text{ is selected for a Biodiversity Management Area} \\ 0 & \text{if not} \end{cases}$

We can formulate the Biodiversity Management Area Selection (BMAS) Model in the following manner:

$$Minimize\ Z = \sum_j \left(w_1 a_j + w_2 Hd_j + w_3 r_j + w_4 Pla_j + w_5 PPI_j\right) X_j \qquad (1)$$

Subject to the following conditions:

1) Element k is sufficiently represented in BMAs to be considered not vulnerable, that is,

$$\sum_j a_{jk} X_j \geq Min_k \ \ for\ each\ k \in K \qquad (2)$$

2) Integer requirements: $X_j = 0$ or 1 for each $j \in J$

The above model deals with biodiversity management area selection (BMAS). This model involves selecting the most suitable and most compatible planning units that overlap with vulnerable elements. Enough area covering each vulnerable element must be selected to remove it from being vulnerable. This is established in the constraints of type (1). Either a planning unit is selected as a BMA or it is not. This is enforced by the definition of the integer decision variables and formalized in constraints (2). The objective function contains a number of terms. The first term is strictly an area term, so there will always be some concern for minimizing the total area selected for biodiversity management options. With the exception of the first term of the objective, each additional term involves a measurement of the suitability and potential effectiveness for that area to be used as a biodiversity management area. For each term, e.g. r_j, a high value reflects low compatibility or potential as a BMA and a low value is indicative of being very suitable and compatible with targeting as a BMA. Specifically, r_j and Hd_j are terms which measure habitat quality, while the terms, Pla_j and PPI_j are terms which reflect the suitability of an area based on management and cost concerns. The objective involves minimizing the total area selected as well as optimizing the suitability of those areas selected by minimizing any such incompatibility.

The purpose of the BMAS model is to help guide in the selection of areas which can be valuable in a core of management areas for biodiversity. The major objective is to identify areas that are both appropriate for the protection and enhancement of biodiversity, but also identify enough area to represent at least a minimum amount of a given element's distribution. This is an important departure from many of the biodiversity -reserve design models that have been developed in the past. For example, many of the reserve design models have an objective of including each element at least once or a prespecified number of times. Picking polygons that include an element does not address whether that element is widely abundant or whether enough area is available within that area alone for biodiversity protection and representation. The use of a model which optimizes the number of distinct areas which include an element is valuable, but without related analyses which involve total area involved or needed by a given element, is limited and can be misleading.

If we change the Min_k values to the number of times an element is to be represented and change the a_{jk} coefficients to represent the presence of an element (or not) in a given planning unit, the above BMAS model can represent reserve design models based on the species covering approach. This means that the BMAS model is a relatively general model construct and can represent reserve design models like that of Margules et al. (1991). Consequently, developing techniques for solving the BMAS model is an improtant research objective.

Unfortunately, the BMAS model is related to the class of n-p hard problems that can be found in the integer programming literature (like the travelling salesman problem). This can be easily demonstrated as the BMAS model can be transformed into an equivalent multi-dimensional knapsack problem. Basically, worst-case instances of large n-p hard problems may require an inordinate amount of computer time to solve optimally. Our use of a general purpose Optimization Subroutine Library (OSL by IBM) to solve moderate sized BMAS problems has been modestly successful at best. Consequently, most of our research has been focused on the design of a robust heuristic to solve the BMAS problem. Our heuristic is based on the combination of several well-known methods including greedy, interchange, and multiple drops and adds (which represents a form of strategic oscillation). The details of the approach are given in Okin et al. (1995). In testing the heuristic against known bounded solutions for selected problems, heuristic performance was consistently within 2% of optimality.

BMAS Model Application

The SNEP area was divided into six separate planning regions whose boundaries were defined by major river drainages. This division was made to capture latitudinal and longitudinal gradients in Sierran habitats (see Davis et al. 1996). Watersheds make logical units for BMAs because they are readily located on the ground, are appropriate physiographic units for managing ecosystem and hydrologic processes, and may be large enough to support viable populations of many plant and animal species. We used the Calwater planning watersheds as our basic BMA unit (which average 3000 ha (7,000 ac) in size). The areal extent of every plant community types in each watershed was calculated by intersecting the watershed boundaries with a map of plant community types (Davis and Stoms 1996). The vegetation map was prepared at 1:100,000 scale for the gap analysis of the Sierra Nevada. In general, there are several community types per watershed.

In each region we defined a starting BMA system based on maps of land ownership and management. (For example, one alternative is to consider all parks, designated nature reserves, and ungrazed designated wilderness areas as BMA lands.) Next we established a target level for representing plant community types in BMAs. This level can be set for each individual element, but for simplicity we used the same target level, for example, 10% of the mapped distribution, for every plant community type. By overlaying the map of existing BMAs on the map of plant community types we determined which types are not adequately represented and how much additional BMA land is needed for each type. This forms the basis of the Min_k values.

Our purpose is not to identify the optimal sites for a Sierra BMA system. The BMAS model allows us to explore some of the likely dimensions of plausible, alternative BMA systems for the Sierra Nevada to answer the set of questions posed above. Alternatives can be generated readily by changing the model inputs: e.g., assumptions about current management, target levels for reducing vulnerability, the land base from which new BMA's can be selected, the weights in the objective function, the suitability factors, or the biodiversity elements to be represented.

Detailed results of the BMAS model applied to the Sierra Nevada can be found in the SNEP final report (Davis et al. 1996). Here we limit our presentation to one example alternative solution of the BMAS model for the northern subregion of the Sierra Nevada (Figure 1). This alternative identifies the most suitable BMA core which contains at least 10% representation of the distribution of each vulnerable element. For this alternative, we assumed that only designated reserves, parks, and wilderness areas which are not grazed are currently managed for biodiversity (approximately 2% of the subregion). With these assumptions and targets, 54 of the 59 plant community types in the subregion would be considered vulnerable. The model selected 55 out of 776 watersheds as new BMAs, totaling 189,000 ha. Private land comprised 41% of the new areas, whereas 48% of the subregion as a whole is privately owned. Thus, the suitability objective in the model, which is weighted against private ownership, helped direct selection towards public lands but because the foothill areas are so predominately in private ownership, several community types could only be found there.

Figure 1: An example BMAS solution for the northern subregion of the Sierra-Nevada Region. Selected watersheds are outlined in red. This alternative was based on a target level of 10% protection for every native plant community type.

The model was also quite efficient in selecting watersheds. The combined area of new and existing BMAs was only 10.8% of the land area. The additional 0.8% beyond the 10% target of this alternative is due to the requirement that entire planning watersheds be selected, even if only part would have been required to meet the Min_k constraint. Furthermore, five community types already exceed the 10% representation requirement and were not considered vulnerable. Note that there is a moderate level of natural clustering of several of the watersheds that are selected by the model. This apparent clustering we attribute to the underlying spatial autocorrelation of both the distributions of plant communities and of the suitability factors, at least at the resolution of the planning watersheds used in the analysis. As the target level is increased to 25% representation, the clumping seems even more pronounced, although we have not formally measured this spatial pattern. If this clumping pattern is real, it suggests that the aggregation of selected watersheds in the 10% solution is limited primarily by the Min_k constraint.

The current version of the BMAS model does not consider the spatial pattern of the selected watersheds. Based on general principles of conservation biology one could argue that larger, better connected BMAs would tend to maintain biodiversity better than small, poorly connected systems (Reid and Murphy 1995). On the other hand, there is evidence that populations in several scattered sites are less vulnerable to large-scale environmental disturbances than populations in a single larger site (Harrison and Quinn 1989). Obviously, it would be useful to incorporate spatial considerations in the BMAS model in order to explore these issues more analytically. Contiguity is difficult to incorporate as a suitability factor, however, because it is not a property than can be measured a priori for a watershed but is dynamic in that it changes as its neighbors are selected. The BMAS model used here provides solutions that are the most efficient solutions only in terms of requiring the least area. Thus the solutions can be considered planning benchmarks in terms of the area requirements for representative BMA systems. Any additional constraints such as spatial design will increase the area of the solution.

Acknowledgements

Financial support for this research was provided by the USFS Sierra Nevada Ecosystem Project. Computing support was provided by a grant from the IBM Corporation Environmental Research Program. We gratefully acknowledge Michael Bueno and Joe Walsh of UCSB for assistance and many individuals from SNEP for supplying data. We particularly appreciate the productive discussions with Norm Johnson which refined the model and the BMAS alternatives considered.

References

Davis, F. W. and Stoms, D. M. (1996) Sierran vegetation: A gap analysis. In *Sierra Nevada Ecosystem Project: Final Report to Congress, vol. II, Assessments and Scientific Basis for Management Options*, in press.

Davis, F. W., Stoms, D. M., Church, R. L., Okin, W. J., and Johnson, N. L. (1996) Selecting biodiversity management areas. In *Sierra Nevada Ecosystem Project: Final Report to Congress, vol. II, Assessments and Scientific Basis for Management Options*, in press.

Harrison, S., and Quinn, J. F.. (1989) Correlated environments and the persistence of metapopulations. *Oikos* 56: 293-298.

Margules, C. R., Pressey, R. L., and Nicholls, A. O. (1991) Selecting nature reserves. Pp. 90-97 in, C. R. Margules and M. P. Austin, eds., *Nature conservation: cost effective biological surveys and data analysis*. Melbourne: CSIRO.

Moyle, P. I., and Randall, P. J. (1996) Biotic integrity of watersheds. In *Sierra Nevada Ecosystem Project: Final Report to Congress, vol. II, Assessments and Scientific Basis for Management Options*, in press.

Okin, W. J. (1996) *Solving the Biodiversity Management Area Selection Problem*, Masters Thesis, University of California Santa Barbara, in progress.

Reid, T. S., and Murphy, D. D. (1995) Providing a regional context for local conservation action. *Bioscience Supplement*: S84-S90.

Author Information

Richard L. Church, Professor
Department of Geography
University of California, Santa Barbara
Santa Barbara, CA 93106-4060
Voice: (805) 893-4217

Fax: (805)-893-3146
e-mail: church@geog.ucsb.edu

David M. Stoms, Assistant Research Scientist
Institute for Computational Earth System Science
University of California, Santa Barbara
Santa Barbara, CA 93106-3060
Voice: (805) 893-7655
Fax: (805)-893-3146
e-mail: stoms@geog.ucsb.edu

Frank W. Davis, Professor
Department of Geography
University of California, Santa Barbara
Santa Barbara, CA 93106-4060
Voice: (805) 893-3438
Fax: (805)-893-3146
e-mail: fd@geog.ucsb.edu

William J. Okin, Graduate Student Researcher
Department of Geography
University of California, Santa Barbara
Santa Barbara, CA 93106-4060
e-mail: okin@geog.ucsb.edu

L. Edward Harvey

Macroecological studies of species composition, habitat and biodiversity using GIS and canonical correspondence analysis

Current ecological research is dominated by attempts to generalize ecological processes and function from controlled, replicated ecological experiments of few species. However, recent developments in macroecology have drawn upon data for many species at regional and continental scales. One approach has been to make ecological infererences from multivariate statistical analyses of species assemblages and environmental conditions. Canonical correspondence analysis (CCA) is one such method commonly used in ecology and biogeography to analyse variation in species composition and environment among samples. Extensions of this robust statistical method to macroecology can provide useful information about geographic variation in community composition. Current spatial applications of CCA are reviewed, and methods for integrating CCA with GIS coverages of plant and animal assemblages are examined.

An analysis of the avian biodiversity of a New Zealand reserve is used to demonstrate methods of combining CCA with GIS in macroecological modelling. CCA is used to produce maps of overall bird habitat quality and heterogeneity. The distributions of several bird species are predicted from CCA scores using generalized additive and classification tree models. CCA can provide useful ecological insights when applied to the spatial scale normally studied by biogeographers. CCA and GIS are a powerful combination to analyse regional patterns of species composition and species-habitat usage, and to assist with mapping predicted species distributions, biodiversity, and habitat heterogeneity.

Introduction

"Macroecology" is a new arena of research that bridges ecology and biogeography by substituting a comparative statistical methodology for microscale manipulative experiments. It synthesizes ecological and biogeographical phenomena, and patterns and processes in other earth sciences to address fundamental questions about relationships between the distribution, abundance and diversity of organisms and the environment at large spatial and temporal scales (Brown and Maurer 1989, Brown 1995). The emergence of macroecology stems mainly from theoretical advances in modelling the complexity and scale-dependence of ecological systems, the availability of large continental databases of species distributions and abundance, and technological advances in compute power and spatial analytical techniques. Microscale detail is sacrificed to reveal larger spatial and temporal scale ecological patterns representing whole-system properties and processes in the structure and dynamics of ecological systems. The breadth of research topics in macroecology and techniques for analyzing the geographical structure of populations are described in Brown (1995) and

Maurer (1995).

At the core of the interest in macroecology is a realization by many biogeographers and ecologists that applied disciplines such as conservation biology and resource management remain focused at fine-scale issues, and their methods are not easily extrapolated to address global ecological problems from intensifying human activity (Brown and Maurer 1989). Scaling up from fine-scale ecological observations can create serious modeling problems because some ecological processes are nonstationary at regional or geographical scales (Hengeveld 1994), and species distributions are primarily an integrated result of recent human impacts on the environment (Miller 1994). Threats to biodiversity from regional habitat loss, degradation, fragmentation and the spread of biological invaders are macroecological. Therefore, geographic models of community composition and biological diversity (or biodiversity) must be based on regional biological surveys and spatially explicit statistical models. Local conservation management for the long-term persistence and ecological integrity (Woodley et al. 1993) of populations, species and ecosystems will require predictions from empirical models based on patterns and processes that operate at landscape scales.

Regional models of species' abundance, distribution and diversity will be most accurate when the macroscope is composed of tools such as remote sensing, geographic information systems (GIS) and techniques for statistical analysis of large-scale spatial databases. GIS has assumed a central role in numerous species-specific applications (e.g. Downley et al. 1992, Duncan et al. 1995, Goldblatt 1993, Pereira and Itami 1991) but there is more scope for GIS in modeling species assemblages, scale-dependent habitat preferences, and macroecological topics such as the geographical fragmentation of populations, habitat heterogeneity, and ecological integrity. Sophisticated multivariate methods for analyzing species composition have been developed in community ecology and recently extended to landscape ecology (Jongman et al. 1995). Gradient analysis is one such family of techniques with relatively untested potential in macroecology. Species distribution data have been used in direct gradient analysis methods to measure the realized climatic niche response of individual plant species (Westman 1991). However, gradient analysis methods have not yet been integrated with GIS for spatial analysis of species distributions and assemblages.

In this paper I focus on the integration of gradient analysis and GIS for mapping species composition across landscapes, constructing empirical predictive models of species distribution, and quantifying species-habitat relationships. This combination of multivariate statistics and GIS has the potential to assess the indicator properties of species assemblages (Kremen 1992), quantify the effects of habitat fragmentation on species composition (Farina 1995, Lescourret and Genard 1994), identify biodiversity hotspots, prioritize areas and activities for conservation, assess the biogeographic representativeness of reserves (Saetersdal and Birks 1993), and identify gaps in conservation networks (Linzey and Harvey 1995).

The main objectives of this paper are to: 1) review recent developments in gradient analysis for modeling geographical species/environment relationships, and 2) evaluate the advantages of combining canonical correspondence analysis (CCA) with GIS for macroecological research. Results from a gradient analysis of geographical patterns of bird species composition in a New Zealand reserve are presented to demonstrate the uses of CCA for modelling species distributions, habitat preferences and biodiversity.

Canonical correspondence analysis (CCA)

Direct gradient (or regression), indirect gradient (ordination), and classification (cluster) families of multivariate statistics are popular in community ecology. Classification techniques are discussed by van Tongeren (1995), and comprehensive reviews of gradient analyses are provided by ter Braak (1995) and ter Braak and Prentice (1988). In general, gradient analyses are multivariate statistical methods designed to analyze biogeographical and ecological data sets of species occurrence and environmental variables over numerous sites. The ecological niche provides a conceptual and empirical basis for gradient analysis; each species is treated as a distinct biological unit with unique ecological requirements that are reflected in its abundance and distribution over space and time (Brown 1995).

The diverse array of gradient methods is summarized in Table 1. Some methods have been designed specifically for species assemblage data to reduce the "noise" of individual species distributions and reveal significant species/environment relations. Most direct gradient methods use iterative optimization techniques to arrange sites along environmental axes based on the species composition and environmental conditions at each site (ter Braak 1987). Nonlinear (unimodal) statistical models are preferable to linear models for species/environment relationships because species abundance usually follows a normal distribution curve along each environmental gradient (ter Braak and Prentice 1988). This is especially the case at a regional and continental scale as species ranges can be considered broad-scale response surfaces with survival probability declining away from a central region where the species' ecological intensity (abundance or presence/absence) is highest. Species distributions have ecological limits defined by an interaction between species' niche requirements and the abiotic and biotic characteristics of the environment. Biotic interactions can be proximate limiting factors, but species distributions are determined ultimately by the physical template (Brown 1995).

Table 1. Summary of gradient analysis techniques classified by type of response model (linear, unimodal) and types of variables. MR, normal multiple regression; IR, inverse regression; PCA principal components analysis; RDA, redundancy analysis; COR, canonical correlation analysis; WAE weighted averaging of environmental variables; GLM, generalized linear modeling; ML, maximum likelihood; WAI, weighted averaging of indicator values; CA, correspondence analysis; DCA, detrended correspondence analysis; CCA, canonical correspondence analysis; DCCA, detrended canonical correspondence analysis (modified from ter Braak 1995).

	Response model		**Number of variables**	
	linear	*unimodal*	*response (species)*	*explanatory (environmental)*
Regression	MR	WAE,GLM,ML	one at a time	>=1*
Calibration	IR	WAI,ML	>=1*	rarely >1

Ordination	PCA	CA,DCA,ML	many	none
Canonical ordination	RDA	CCA,DCCA,ML	many*	many*
	COR	variants of CCA,ML	many*	many*

* less than the number of sites, except for WAE, WAI and some applications of ML.

Canonical correspondence analysis (CCA) is currently the most advanced and robust gradient analysis method (Palmer 1993). Its basic aim is to derive from a linear combination of environmental variables a series of canonical axes that are restricted or "constrained" to be weighted sums of the environmental variables. Goodness-of-fit is calculated as the ratio of the within-species sum-of-squares of the canonical axis to the overall sum of squares (Hill 1991). The importance of each CCA axis is represented by the eigenvalue, which measures how much variation in the species data is explained by the combination of environmental variables for the axis (ter Braak 1995).

CCA is based on the assumption that the species response distribution is unimodal, as represented by Gaussian response curve,

$$z = c \cdot \exp\left[-0.5(x-u)^2 / t^2\right] \quad (1)$$

where *z* is the original abundance value, *c* is the species' maximum abundance, *u* is its optimum (the value of *x* that gives maximum abundance), and *t* is its tolerance (a measure of ecological intensity) (ter Braak and Looman 1995). CANOCO ver. 3.12 (ter Braak 1990) calculates the species score as the center of this distribution (*u*) and the width is quantified by the standard deviation (*t*) or tolerance, which is a measure of niche width. The site score indicates the center of the site response distribution and is restricted to be a linear combination of the measured environmental variables in CCA (ter Braak 1995). The standard deviation is used to measure site heterogeneity, which represents the variation in species composition, abundance, and environment of each site. Mapping this heterogeneity statistic may reveal spatial patterns of the overall biodiversity (i.e. species composition and habitat) for the study taxa, and thereby illustrate the enormous spatial heterogeneity in abundance that is characteristic of species distributions (Brown 1995).

Gradient analysis results are most often displayed in diagrams where sites are represented by points and their relative position reflects their similarity in species composition. The diagrams are a graphical summary of the variation in abundance among species and species composition among sites with respect to the canonical axes, and are valuable for interpreting complex interrelationships between species composition and environment. However, results from gradient analyses can also be used in other multivariate analyses and GIS to model species distributions and map spatial patterns of species composition. Some of these applications are discussed below.

Macroecological applications of CCA

Predicting species occurrence from canonical axes

Predicting the distribution of a species requires good survey data as well as knowledge of environmental factors and autoecology (Le Duc et al. 1992). However, populations are rarely mapped because the process is labor-intensive, most populations have dynamic distributions, and most remote sensing techniques cannot detect small organisms (Johnston 1993). As a result, spatial data of species abundance are available mainly for a few endangered or threatened species. Atlases of plant and animal distributions contain basic spatial data on the presence/absence of species, but these simple maps are gaining importance in biological conservation for predicting the effects of environmental degradation and climate change (Smith 1994). Equal-area grid atlases based on systematic regional surveys are also easily converted to a raster GIS. The raster is a flexible structure for data storage, sampling, cross-validation, spatial analysis, modeling, and display. McAllister et al. (1994) demonstrate the usefulness of equal-area grids and GIS for analyzing global patterns of coral reef fish diversity. However, species richness is a scale-dependent ecological phenomenon so maps of species richness are sensitive to habitat map grid (or minimum mapping unit) size (Stoms 1992, 1994).

Biological atlas data can be used to model the geographical distribution of individual species from climatic, geological, edaphic and physiographic data, and their probabilities of occurrence predicted for unsampled areas. Predicted distribution maps can then be combined to provide species lists for a given area and map regional patterns of species richness. Logistic regression models are commonly used to quantify species-specific habitat preferences (or tolerances), and map probabilities of species occurrence. For example, Gates et al. (1994) used logistic regression to model British bird species occurrence in 10 km-square grid cells from several environmental variables. Similarly, Osborne and Tigar (1992) used bird atlas data and principal components of habitat variables in a logistic regression to predict bird species distributions in Lesotho, southern Africa.

The logistic regression model is a generalized linear model (GLM) specifically designed for modeling binomial data (Hastie and Pregibon 1993). It uses a linear combination of independent variables (i.e. canonical axes or original habitat variables) to explain the variance in a dependent variable with two states (i.e. species presence/absence). The assumption that a species' occurrence relates to an environmental gradient in a logistic rather than a linear manner is consistent with ecological theory that prescribes a sigmoid-type curve for species tolerance over part of the gradient (Osborne and Tigar 1992).

Using canonical axes in spatial logistic regression models

The distribution of many species is best explained by their relationship with environmental gradients, which can be represented as canonical axes. This approach has been used to analyze biogeographic patterns of many taxa. Owen (1990) modeled Texan mammal species distributions and their environmental relationships. Hill (1991) used gradient analysis in combination with logistic regression to predict occurrence of bird and plant species in km^2 squares in Britain. Similarly, Carey et al. (1995) conducted a detrended canonical correspondence analysis (DCCA) of six taxonomic groups for 10 km-square grid cells in Scotland. The first four canonical axes were used in a *k*-means cluster analysis to identify natural biogeographic zones. Carabid beetles have been popular taxa for gradient analysis of species distributions and habitat preferences. For example, Dufrene (1990) conducted a *k*-means cluster analysis of PCA axes created from carabid beetle species presence/absence data for 10 km^2 grid squares in Belgium, and Heliovaara et al. (1991) analyzed biogeographic patterns of carabids in Fennoscandia and Denmark using gradient and cluster methods.

However, positive spatial autocorrelation in the probabilities of species occurrence violates the assumption of independent observations in regression models. Various methods have been used to include the spatial structure of species distributions and environmental variables in predictive models. Some have exploited underlying spatial autocorrelation with geostatistics to improve the fit of predictive models (Liebhold et al. 1993, Smith 1994). For example, Le Duc et al. (1992) used a spatial Gaussian smoothing technique in combination with logistic regression to estimate individual response surfaces for plant species in Britain. An alternative strategy is to remove spatial structure from the predictive models. The former strategy generally produces more accurate predictions of species occurrence within the study region, whereas the latter strategy produces species habitat models that may be sufficiently general to predict species occurrence outside the sample region.

Methods of exploiting spatial autocorrelation are most useful for optimizing predictive models of species occurrence when model complexity is not an issue. However, when the objective is to reveal species-specific habitat preferences then a more appropriate research strategy is to first remove underlying spatial structure from the biological and environmental data, and then model the residual species-environment relationships. Furthermore, over regional scales species composition and environmental variables may share a common spatial structure. In this situation the spatial structure will cause gradient analyses like CCA to overestimate the amount of species variation (sum of canonical eigenvalues) explained by the environmental variables. Recent developments in gradient analysis have focused on methods for partitioning variation in species composition among sites into two or more components. Borcard et al. (1992) demonstrate how this partitioning approach can be employed with CCA to quantify species-environment relationships after removing the underlying spatial structure. They use four CCA analyses to partition the species variance :

1. CCA of the species matrix, constrained by the environmental matrix,
2. CCA of the species matrix, constrained by the matrix of spatial coordinates,
3. like (1), after removing the effect of the spatial matrix,
4. like (2), after removing the effect of the environmental variables.

Eigenvalues from these four steps are used to calculate the overall amount of explained variation (step (1) + step (4), or step (2) + step (3)), the relative percentages of variance explained at each step, and partition the total species variance:

- nonspatial environmental variation = step (3),
- spatially structured environmental variation = step (1) - step (3), or step (2) - step (4),
- spatial species variation not shared by the environmental variation = step (4),
- unexplained variation and stochastic fluctuations = 100 - (step (1) + step (4)).

Partitioning has been used successfully (e.g. Okland and Eilertsen 1994, Vetaas 1993) to more accurately infer species-specific habitat preferences, independent of underlying spatial patterns in both species and environmental data. However, CCA cannot be used to quantify spatial autocorrelation. Alternatively, Sanderson et al. (1995) use partial Mantel tests to quantify and control for spatial autocorrelation in ordination analyses of species-specific habitat preferences.

Quantifying habitat preferences from gradient analysis statistics

An extensive literature exists for modeling species habitat relationships (see Alldredge and Ratti 1992, Morrison et al. 1992, Porter and Church 1987, Thomas and Taylor 1990). Knowledge of species habitat preferences are a fundamental component of GIS biological gap analyses (e.g. Davis 1994). However, it is often difficult to construct a species-habitat relationships model because the habitat preferences of most species is unknown, and when this knowledge is available it is rarely as detailed as the vegetation maps (Cassidy et al. 1993). Direct gradient analysis methods such as canonical correspondence analysis (CCA) may be useful to infer species-specific habitat preferences from statistical analysis of geographic distribution data and coarse-scale environmental variables (e.g. maps of vegetation, elevation, climate, geology). The basic principle is that spatial variation in the abundance (or presence/absence) of species' among sample sites is a function of the unique habitat preferences of each species and the unique environmental conditions of each site. CCA has the advantage of analyzing simultaneously all the environmental variables potentially determining species habitat selection (Baguette 1993). This multivariate approach is also well suited to model habitat relationships from species assemblage data at landscape and continental scales. It must be emphasized, however, that this empirical statistical approach cannot replace the more accurate species-specific inferences from field surveys and experimental analyses of habitat selection.

Recent examples of gradient analyses of species' habitat at community scales include correspondence analysis (CA) of habitat selection patterns of the woodland carabid fauna of southern Belgium (Baguette 1993), and CCA of seasonal patterns of avian abundance in northeastern Venezuela (Poulin et al. 1993). Similar recent biogeographical habitat modeling applications of gradient analyses include redundancy analysis (RDA) of breeding bird habitat associations in the eastern Highlands of Scotland (Brown and Stillman 1993), CCA of bird species habitat in a Finnish archipelago (Martin and Lepart 1989), and CCA of the habitat preferences of butterflies in northern Vietnam (Spitzer et al. 1993). In each of these studies gradient analyses quantified the relative importance of habitat variables for individual species or groups of species, and identified the dominant environmental gradients responsible for variation in species composition among the study sites.

Another direction of research has attempted to use results from gradient analyses (mainly species and site scores) to construct general models of species habitat preferences and predict species occurrence in unsampled areas (e.g. Osborne and Tigar 1992). These studies commonly use gradient analyses to reduce a large set of environmental variables to a few orthogonal canonical axes. The axes are then used as independent variables in logistic regression models to predict the occurrence of a given species at a given site.

Buckland and Elston (1993) provide a useful empirical modeling framework for the spatial distribution of wildlife. To date, most attempts to model wildlife distributions have used generalized linear models. The presence/absence or a measure of relative or absolute abundance of a species is the dependent variable and a set of environmental variables form the independent variables. Generalized linear models have the advantages of known statistical properties, choice of appropriate link function and error distributions, and they can reflect basic GIS masking and overlay operations (Buckland and Elston 1993). The family of generalized additive models provide an even wider range of modeling options (Hastie 1993).

Some of the most sophisticated and accurate species habitat models for large regions have combined field surveys with remotely sensed data in GIS (Griffiths et al. 1993, Hunsaker et al. 1993, Miller 1994). For example, Duncan et al. (1995) tested a Florida scrub jay habitat suitability index model originally developed from remotely sensed data and field derived demographic data. Rappole et al. (1994) used TM imagery to assess the amounts and rates of change of winter habitat of migratory birds in Costa Rica. Spectral bands, vegetation indices or other land-cover images used to map wildlife habitat could also be used as environmental variables in gradient analyses to supplement more detailed species-specific habitat information provided by measurements of environmental variables at field sites.

Modeling bird species composition in a New Zealand reserve

Conservation efforts in New Zealand have traditionally focused on rare species management, with specific concern for birds (Atkinson 1994). As a result, much of New Zealand vegetation science in protected areas has been in the context of bird habitat (Ogden 1985). To demonstrate some uses for CCA in modeling regional species composition and individual species distributions I present a brief summary of results from ongoing analyses of bird species in the Waipoua Forest Sanctuary, North Island, New Zealand (35 deg. 38' S, 173 deg. 34' E). The Waipoua Forest combined with the contiguous Mataraua and Waima State Forests forms the largest continuous tract of indigenous vegetation in Northland. The indigenous vegetation is composed of 12,851 ha of species-rich podocarp-broadleaf forest, with large remnants of mature kauri (Figure 1). Exotic pine plantations separate most of the Waipoua Forest from the coast. Several gradients in climate, soils and vegetation physiognomy occur over the sea level to 519 m elevation range.

Figure 1. Waipoua Forest Sanctuary vegetation (Department of Conservation 1987). Vegetation codes are in Table 2.

The Waipoua Forest is also habitat for 59 bird species, of which 22 are native, 4 are migratory, 19 are introduced, and 14 are self-introduced (Eadie et al. 1987). The Waipoua Forest is situated in a mostly agricultural landscape mosaic; adjacent developed land is mostly pasture (36%) and plantation forestry (11%) (Eadie et al. 1987). Many native bird species occur in remnants of indigenous vegetation throughout Northland, but with the exception of a few rare or endangered species their distributions have not been mapped (although a national bird atlas is available for a 10,000 yard-square equal-area grid (Bull et al. 1985)) and species-specific surveys of bird habitat quality (Ogle 1981) are not available. As a result, it is difficult to determine the effects of habitat fragementation, or use gap analyses to assess the effectiveness of the reserve network for protecting bird species diversity (Linzey and Harvey 1995).

As an alternative to conducting a systematic regional bird inventory for the Northland region, bird species distributions can be modeled from available bird surveys and data on vegetation, climate, topography, soils, geology and satellite imagery. The models can then be used to predict bird species occurrence in areas of Northland without bird surveys. In this simple case study I demonstrate methods for modeling bird species distributions from existing bird and vegetation surveys in the Waipoua Forest. It is important to test whether CCA can detect known bird species habitat preferences from these spatially coarse data of simple bird presence/absence and general canopy vegetation classes.

Data and General Methods

A 900m X 900m raster of forest bird species data was created from grid maps of species presence/absence for the Waipoua Forest published in Eadie et al. (1987). Twenty-five bird species were recorded in 173 grid cells. The presence/absence of nineteen vegetation types (Table 2) in each grid cell was created from a canopy vegetation map of the Waipoua forest (Department of Conservation 1987) (Figure 1). The average elevation of each grid cell was estimated from the contour line (40 m intervals) closest to the grid cell center. Three data matrices were used in the gradient analyses:

- bird species presence/absence for each grid cell (25 species X 173 cells),
- environmental variables for each grid cell (21 variables X 173 cells), and
- geographical coordinate variables for each grid cell (3 variables X 173 cells).

CCA methods

The program CANOCO version 3.12 (ter Braak 1987, 1990) was used to conduct all gradient analyses and diagnostic statistics. Two stages of ordination analysis were used to relate variation in the bird species composition to variation in the habitat (environmental) variables:

1. Detrended Correspondence Analysis (DCA) was used to determine lengths of the gradients (axes) (Eilertsen et al. 1990). Gradients were sufficiently long (> 2 s.d.) to justify use of CCA, which assumes species have a unimodal response to the environmental gradients (ter Braak 1995) .
2. A series of four CCA analyses were conducted to partition spatial and environmental variation in the bird species data.

CCA canonical axes with spatial structure removed were used to: a) identify the environmental variables which account for most of the variation in bird species composition, b) identify bird species habitat (vegetation) preferences, c) map patterns of habitat heterogeneity and overall biodiversity, d) predict bird species distributions with logistic regression and classification tree models.

Table 2. Canopy vegetation classes in the Waipoua Forest Sanctuary (Department of Conservation 1987).

code	vegetation
F1	Mamangi-mapou-kanuka forest

F2	Taraire/kohekohe-nikau-karaka forest
F3	Kauri/taraire-towai/kohekohe forest
F4	Taraire/kohekohe forest
F5	Rimu-nothern rata/taraire-towai forest
F6	Rimu/towai-taraire-makamaka forest
F7	Rimu/towai-makamaka-pukatea forest
F8	Rimu/towai-taraire-tawa/kiekie forest
F9	Taraire-towai-miro/kiekie forest
F10	Kauri/taraire forest
F11	Kauri/Hall's totara/tawherowhero/kauri grass
F12	Kauri-tanekaha/mamangi-kanuka-towai forest
Scrub	Towai-manuka/kiokio-bracken scrub
Shrub1	Manuka-dracophyllum/gleichenia fern-shrub
Shrub2	Manuka-bracken shrubland
Tussock	Kauri/manuka/gahnia shrub-tussockland
Fern	Baumea-glechenia-manuka sedge fernland
Elevation	Elevation (meters)
P_forest	Production forest, *Pinus radiata*
Unclass	Unclassified vegetation

CCA Results

A forward stepwise regression method was conducted using the CANOCO (ter Braak 1992) gradient analysis software to select the most statistically significant spatial variables (α= 0.05) from a cubic trend surface equation. Details of the spatial partitioning procedure and all CCA results are given in Linzey and Harvey (1995). The spatial structure of bird species composition in the Waipoua forest was modeled as:

$$z = \beta 2\ (E) + \beta 5\ (E\hat{}2) + \beta 9\ (E\hat{}3), \quad (2)$$

where z is species occurrence, E is the easting, and βi are the regression coefficients. The sum of all eigenvalues in the CCA analyses is 1.975, and the overall amount of variance explained is ~23%, which is partitioned as:

- nonspatial environmental variation = 15.1%
- spatially structured environmental variation = 6.5%
- spatial species variation not shared by the environmental variation = 1.3%
- unexplained variation and stochastic fluctuations = 77.1%

Bird species composition varies with spatial location, but most of this spatial dependency is also shared by the environmental variables; only 1.3% of the variance is explained by spatial structure unique to the bird species distributions. Over 77% of the species variance is not explained by these environmental variables. It is common for the unexplained portion of variance to be high in ecological data (ter Braak 1992), but these results are intended only to demonstrate a macroecological application of gradient analysis and GIS.

The gradient represented by each canonical axis is inferred from correlation and regression/canonical coefficients for each environmental variable (see Linzey and Harvey 1995). The first canonical axis represents a vegetation gradient composed of F2 (tararire/kohekohe-nikau-karaka forest) and shrub2 (manuka-bracken shrubland). The second axis represents an elevation gradient from production forest (*Pinus radiata*) to indigenous forest. Vegetation varies mainly from shrubs in the west, and mixed forests of taraire, kauri, rimu or towai in the east. This pattern reflects east-west topographic climatic gradients inland from the coast to the west.

CCA species scores summarize the relative influence of all the environmental variables, and thereby provide quantitative descriptions of habitat preferences. Most bird species have low scores for the first two ordination axes, reflecting their general habitat requirements and widespread distribution. Other species have scores that their reflect specific habitat requirements or rarity. For example, New Zealand pipits (*Anthus novaeseelandiae novoseelandiae*) have high scores for axes 1 and 2, indicating preferences for tararire/kohekohe-nikau-karaka forest, manuka-bracken shrubland and higher elevations, and negative association with exotic forest.

These gradient analysis results suggest that the Waipoua forest is a relatively homogeneous landscape for many of these bird species. However, sites at the periphery of the forest to the west have more unique environmental characteristics and bird species with narrower habitat preferences. Western sites also have the greatest human modification and mix of coastal vegetation. These results also suggest that bird habitat preferences are defined more by vegetation structure than plant species composition. The dominant variation in bird species composition is along gradients of shrub to forest, and indigenous to exotic forest.

Heterogeneity as a measure of bird species and habitat diversity

Spatial patterns in the overall diversity of bird species and habitat are also evident in images of site heterogeneity (s.d. of the CCA site scores) for axis 1, root mean square standard deviation across the four axes (RMSTOL), and species richness (N2) (Figure 2). Heterogeneity in this context represents the overall biodiversity (species abundance, composition, and habitat) for birds in the Waipoua Forest. Sites near the coast in the northwest have the highest first canonical axis heterogeneity (Figure 2a), and moderately high heterogeneity over all axes (Figure 2b). Species richness (Figure 2c) fails to identify the high bird species and habitat biodiversity in the northwest.

Figure 2. Heterogeneity of bird species biodiversity in the Waipoua Forest: a) standard deviation of site scores for CCA axis 1, b) root mean square standard deviation across the four CCA axes (RMSTOL), and c) species richness (N2).

Species distribution modeling

Species distributions were predicted from bird species presence/absence data and CCA site scores for the first four canonical axes using logistic regression models. First a stepwise generalized additive method (GAM) was used in a screening procedure to identify the subset of linear or smoothed canonical axes which best fit the distribution of each species. GAM diagnostic graphics were used to determine the appropriate transformation to approximate smoothed canonical axes. The subset of linear and smoothed canonical axes for each species was used in a generalized linear model (GLM) to predict their distribution. Coefficients from the GLM are estimated for a parametric relationship, unlike those for additive models (Hastie 1993).

The best fitting logistic regression models are given for three representative bird species in

Table 3. Chaffinch (*Fringilla coelebs gengleri*) are an introduced granivorous species common in a variety of forest and shrub habitats, New Zealand kingfishers (*Halcyon sancta vagans*) are an indigenous mostly insectivorous species preferring open forests and shrubland, and New Zealand pigeons (*Hemiphaga novaeseelandiae novoseelandiae*) are an indigenous frugivorous species restricted mainly to native forest remnants. Bivariate logistic surfaces representing the best fitting models are shown in Figure 3. Waipoua vegetation gradients are relatively short at the spatial scale and minimum mapping unit used here. This has the effect of producing logistic planes for chaffinch and kingfisher (Figures 3a,b), suggesting that the range of habitats in the Waipoua Forest is small and perhaps suboptimal. The slightly more complex logistic surface for pigeons (Figures 3c,d) illustrates the nonlinear nature of interacting explanatory (habitat) variables.

Figure 3. Bivariate logistic surfaces representing the best fitting logistic regression models (Table 3) for: a) chaffinch, b) NZ kingfishers, c,d) NZ pigeons. The two explanatory canonical axes are plotted in the horizontal plane and black squares represent grid cells where each species was observed. The probability of species' occurrence is shown by lines of equal probability (a c), and as the vertical dimension of a three-dimensional plot (d).

Table 3. Regression coefficients in the best-fitting logistic regression model to the presence/absence of bird species in the 173 900-m grid cells in the Waipoua Forest. Standard errors are in parentheses. For NZ pigeon the $\beta 4$ coefficient is for $(axis4)^2$. Prediction equation:

$$y = \frac{\exp(z)}{1 + \exp(z)} \quad (3)$$

where $z = \alpha + \beta 1\,(axis1) + \beta 2\,(axis2) + \beta 3\,(axis3) + \beta 4\,(axis4)$

Coefficients

	α	$\beta 1$	$\beta 2$	$\beta 3$	$\beta 4$	residual deviance	df	pseudo r^2
chaffinch	0.801	-.320	-	-0.618	-	190.3857	160	0.920
	(0.179)	0.166		(0.181)				
NZ kingfisher	-0.058	-0.382	0.675	0.427	-	200.958	159	0.891
	(0.169)	(0.173)	(0.183)	(0.171)				
NZ pigeon	-0.154	-	-	-0.431	-4.167	213.575	159	0.948
	(0.163)			(0.163)	(2.496)			

Bird distributions were also predicted from canonical axes using a tree-based regression model. The classification tree provides a probability model for the occurrence of a bird species at each grid cell. Computationally intensive binary recursive partitioning algorithms are used to successively split the data into increasingly homogenous subsets. A measure of the cross-validated predictive power of a given split weighed against the complexity of adding the split is used to stop the subdivision and determine the size and structure of a binary tree model (Hollander et al. 1994). Clark and Pregibon (1993) provide a good overview of tree-based models.

Classification and regression tree models (Clark and Pregibon 1993) have been applied in ecological land classifications based on remotely sensed data (Davis and Dozier 1990, Michaelsen et al. 1994) but are a relatively recent development in species distribution modeling. Hollander et al. (1994) used a climate-based hierarchical classification tree model to predict orange-throated whiptail (*Cnemidophorus hyperythrus*) distribution in southern California and Baja California. They suggest that logical formulations of rule-based models may be more appropriate for conservation planning than gradient models.

The classification tree models for chaffinch and kingfisher were pruned from 24 to 8 and 17 terminal nodes respectively on the basis of crossvalidation diagnostics. This procedure is analogous to variable selection in regression (Venables and Ripley 1994). The final pruned classification tree model for chaffinch used CCA axes 1, 2, and 3, with a 17.8% misclassification error rate. All four CCA axes were used in the pruned classification tree model for kingfisher, but misclassification was higher (20.86%).

The above logistic regression and pruned classification tree models were used to predict the distribution of chaffinch (Figure 4a) and kingfisher (Figure 5a) in the Waipoua Forest. Proportional grid squares illustrate the predicted probability of chaffinch occurrence (for $P > 0.5$) from logistic regression (Figure 4b) and classification tree models (Figure 4e). The predicted distribution of NZ kingfishers is shown in Figures 5b, 5e. The accuracy of predicted distributions is artificially high because all models were derived from the full sample of grid sites. However, patterns of two types of model error (deviance residuals) are informative: failure to predict occurrence ($P < 0.5$) where the species was observed, and predicted occurrence ($P > 0.5$) in grid cells where the species was not observed. Buckland and Elston (1993) discuss use of these errors for inferring spatial patterns of habitat suitability. Classification tree models may be more effective at predicting occurrence where the birds were actually observed, whereas the linear logistic regression models tend to predict occurrence in more grid cells where the birds were not observed. The latter could be true model error (i.e. the birds are absent from the predicted cells) or it could represent grid cells where the birds occur but observers failed to observe them. A thorough comparative evaluation of logistic and classification models and their errors is necessary before they can be used to direct future biological field surveys.

Figure 4. Maps of chaffinch observed occurrence (a), and proportional grid squares of the probability of chaffinch occurrence (for $P > 0.5$) predicted by logistic regression (b-d) and classification tree (e-g) models. Two types of model error (deviance residuals) are illustrated: failure to predict occurrence ($P < 0.5$) in grid cells where the species was observed (c,f), and predicted occurrence ($P > 0.5$) where the species was not observed (d,g).

Figure 5. Maps of kingfisher observed occurrence (a), and proportional grid squares of the probability of chaffinch occurrence (for $P > 0.5$) predicted by logistic regression (b-d) and

classification tree (e-g) models. Two types of model error (deviance residuals) are illustrated: failure to predict occurrence ($P < 0.5$) in grid cells where the species was observed (c,f), and predicted occurrence ($P > 0.5$) where the species was not observed (d,g).

Where is the macroecological niche for gradient analysis and GIS?

There are few biogeographical applications that combine gradient analyses and GIS. Cherrill et al. (1995) used detrended correspondence analysis to select a subset of plant community variables for use in a GIS to predict the distribution of individual plant species. Clark et al. (1993) used a simple multivariate statistic in a GIS to develop a habitat model for black bears. Although gradient analyses were not used, their work is significant for integrating habitat modeling in a GIS and recognizing the inadequacies of simple univariate statistical techniques to assess the multidimensional nature of habitats. Some of the procedures for modeling wildlife distributions with GIS discussed by Buckland and Elston (1993) could be extended to include gradient analysis.

Direct gradient analysis methods such as CCA offer some unique benefits for modeling species assemblages. *A priori* knowledge of species' habitat preferences will help to choose the most appropriate environmental variables for modeling their distributions, but the habitat preferences of many species are unknown, and many environmental variables are highly correlated. In the latter case empirical models will agree with closely with the observations, but will give poor predictions when extrapolated to unsurveyed sites (Buckland and Elston 1993). As with other multivariate data reduction techniques, gradient analysis has the advantage of objectively reducing the number of potentially relevant environmental variables.

Other benefits of employing gradient models with GIS in macroecology include:

- applicable to species presence/absence or abundance data, and categorical or numerical environmental data.
- effective for the sparse matrices (with many zeros) common for species X site data.
- reduce the number of independent environmental variables before modeling. This is useful when the number of variables is large because many are irrelevant for explaining species composition; CCA focuses on those variables that are most relevant to the complex gradients. Furthermore, many environmental variables are also highly correlated. Models based on these variables may agree closely with the observations but give poor predictions when extrapolated to unsurveyed sites (Buckland and Elston 1993).
- robust to the statistical effects of rare or abundant species.
- the linear combination of environmental variables represented by each canonical axis may be more realistic representations of the complex environmental conditions selected by each species than the original independent environmental variables.
- the unimodal Gaussian model of the species-environment relationship on which CCA is based has a strong theoretical foundation in ecology.
- ability to partition the effects of spatial structure and subsets of environmental variables (e.g. vegetation, topography, human disturbance, landscape structure).

However, several limitations must also be considered:

- software links between GIS and gradient analyses are currently not available. Data and results must be exported in a common format from independent gradient analysis software (e.g. CANOCO) and GIS. New modules that efficiently link gradient analysis software with GIS to effectively exploit the spatial database will need to be developed. The demand is probably insufficient to expect development of a fully integrated gradient analysis in GIS.
- gradient methods are robust (Palmer 1993), but they are most effective for simultaneous analysis of an assemblage of species at many sites. When data for only a single-species are available, tree-based, generalized linear and generalized additive models may be more appropriate for directly modeling their distribution and environmental response curves.
- the unimodal distribution on which CCA is based may not be appropriate for the response curves of some species (Austin and Gaywood 1994, Austin et al. 1994) because the shape of the curve will be influenced by competition from the assemblage of species competitors that exist over the environmental gradients (Westman 1991).
- large sample sizes from the extent of the species range are required to ensure that the response curve represents the high geographic variability of species performance (Westman 1991).
- habitat selection is highly scale-dependent and species-specific but these are difficult to accommodate in gradient analyses.
- multivariate methods like CCA may suggest species habitat preferences that are inaccurate, or statistical artifacts.
- integrating gradient analyses and multivariate modeling techniques with GIS requires large regional data bases and the analyses are computationally and technologically intensive.

Conclusion

Hollander et al. (1994) suggest that the most fruitful approach to infer habitat preferences and predict species distributions is to combine results from analyses of a variety of data sets and modeling approaches in an interactive GIS environment for spatial analysis and mapping. Integrating ecologically-based gradient models of species-environment relationships (such as CCA) with generalized linear or tree-based statistical models is a potentially useful avenue for modelling in macroecology. This synthesis can suggest habitat preferences for species for which habitat requirements are poorly known and quantify the environmental tolerances and response curves for species whose habitat preferences are hypothesized. Results are also relatively easy to employ in a GIS to map habitat heterogeneity and probabilities of individual species occurrence. CCA cannot provide the detailed knowledge of habitat usage gained by field studies, but it is useful for identifying general species-habitat relationships in the absence of detailed field data. Gradient analyses are especially valuable for generating hypotheses about species/habitat relationships that can then be tested in the field.

Integrating CCA methods with GIS also provides macroecological tools for identifying hotspots of biodiversity. Canonical axes can be used in a variety of regression models to predict individual species distributions, which are easily overlaid in a GIS to identify hotspots of species richness. However, gradient analyses can also be used in more sophisticated biodiversity models to reflect the fact that hotspots are a manifestation of species interactions, ecological and biogeographical processes which vary over space and among species, and historical patterns of human activity (Miller 1994). Direct gradient analyses are specifically designed to quantify the relationships between species abundance and environment that form the ecological base of geographical patterns of biodiversity. Integrating gradient analyses with GIS may provide insights into the mechanisms responsible for macroecological patterns of biodiversity, insights that cannot be obtained from controlled manipulative experiments.

Acknowledgments

L.E. Harvey was funded by The University of Auckland Research Committee and the New Zealand Lottery Grants Board. Amelia Linzey provided bird species data obtained from Waipoua forest bird species maps (Department of Conservation). Alex Lee prepared the Waipoua vegetation map from data provided by Manaaki Whenua/Landcare Research, NZ, and the Regional Resource Evaluation Project (Chris Cocklin and Warren Moran, University of Auckland). The RREP is funded by The University of Auckland Research Committee, the New Zealand Lottery Grants Board, and the New Zealand Vice Chancellors' Committee.

References

Alldredge, J.R., and Ratti, J.T. (1992) Further comparison of some statistical techniques for analysis of resource selection, *Journal of Wildlife Management* 56: 1-9.

Atkinson, I. A. E. (1994) Ecological measures for conserving terrestrial biodiversity: a New Zealand perspective, in *Systematics and Conservation Evaluation*, ed. by P. L. Forey, C. J. Humphries and R. I. Vane-Wright, Oxford: Clarendon, pp. 63-79.

Austin, M.P., and Gaywood, M.J. (1994) Current problems of environmental gradients and specis response curves in relation to continuum theory, *Journal of Vegetation Science* 5: 473 - 482.

Austin, M.P., Nicholls, A.O., Doherty, M.D., and Meyers, J.A. (1994) Determining species response functions to an environmental gradient by means of a -function, *Journal of Vegetation Science* 5: 215 - 228.

Baguette, M. (1993) Habitat selection of carabid beetle in deciduous woodlands of southern Belgium, *Pedobiologia* 37: 365-378.

Borcard D., Legendre, P., and Drapeau, P. (1992) Partialling out the spatial component of ecological variation, *Ecology* 73: 1045-1055.

Brown, A.F., and Stillman, R.A. (1993) Bird-habitat associations in the eastern Highlands of Scotland, *Journal of Applied Ecology* 30: 31-42.

Brown, J.H. (1995) *Macroecology*, Chicago: University of Chicago.

Brown, J.H., and Maurer, B.A. (1989) Macroecology: The division of food and space among species on continents, *Science* 243: 1145-1150.

Buckland, S.T., and Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife, *Journal of Applied Ecology* 30: 478-495.

Bull, P.C., Gaze, P.D., and Robertson, C.J.R. (1985) *The Atlas of Bird Distribution in New Zealand*, Wellington: Ornithological Society of NZ.

Cassidy, K., Garton, E. O., Krohn, W. B., Mills, L. S., Scott, J. M., Williams, K., and Csuti, B. (1994) Assessing the predictive ability of gap analysis vertebrate distribution maps, in *How-to Manual for Gap Analysis*, http://www.nr.usu.edu/gap/valvert.html

Cherrill, A.J., McClean, C. Watson, P., Tucker, K. Rushton, S.P., and Sanderson, R. (1995) Predicting the distributions of plant species at the regional scale: a hierarchical matrix model, *Landscape Ecology* 10: 197-207.

Clark, J.D., Dunn, J.E., and Smith, K.G. (1993) A multivariate model of female black bear habitat use for a geographic information system, *Journal of Wildlife Management* 57: 519-526.

Clark, L.A., and Pregibon, D. (1993) Tree-based models, in *Statistical Models in S*, ed. by J.M. Chambers and T.J. Hastie, London: Chapman and Hall, pp. 377-419.

Davis, F.W. (1994) *Gap Analysis of the Southwestern California Region*, NCGIA Technical Report 94-4, Santa Barbara: NCGIA.

Davis, F.W., and Dozier, J. (1990) Information analysis of a spatial database for ecological land classification, *Photogrammetric Engineering and Remote Sensing* 56: 605-613.

Department of Conservation (1987) *Vegetation Map Of Waipoua Forest And Related Environs*, Department of Conservation New Zealand.

Downley I., Pauknerovs, E., Petch, J., Brokes, P., and Corlyon, A. (1992) Habitat analysis and modelling for endangered species, a sample case - black grouse in Zd'arske vrchy, Czechoslovakia, in *Science and the Management of Protected Areas*, ed. by J.H.M. Willison, S. Bondrup-Nielsen, C. Drysdale, T.B. Herman N.W.P. Munro, T.L. Pollock, Amsterdam: Elsevier, pp 271-276.

Dufrene, M. (1990) Zoogeographical analysis of carabid beetles in Belgium, in *The Role of Ground Beetles in Ecological and Environmental Studies*, ed. by N. Stork, Andover: Intercept, pp. 383-388.

Duncan, B.W., Breininger, D.R., Schmalzer, P.A., and Larson, V.L. (1995) Validating a Florida Scrub Jay habitat suitability model, using demography data on Kennedy Space Center, *Photogrammetric Engineering and Remote Sensing* 61: 1361-1370.

Eadie F., Burns, B., and Leathwick, J. (1987) *Waipoua Ecological Survey 1984-1985*, Department of Conservation and Forest Research Institute, New Zealand.

Eilertsen, O., Okland, R.H., Okland, T., and Pedersen, O. (1990) Data manipulation and gradient length estimation in DCA ordination, *Journal of Vegetation Science* 1: 261-270.

Farina, A. (1995) Distribution and dynamics of birds in a rural sub-Mediterranean landscape, *Landscape and Urban Planning* 31: 269-280.

Gates, S., Gibbons, D. W., Lack, P.C., and Fuller, R.J. (1994) Declining farmland bird species: modelling geographical patterns of abundance in Britain, in *Large-Scale Ecology and Conservation Biology*, ed. by P.J. Edwards, R.M. May and N.R. Webb, Oxford: Blackwell.

Goldblatt, I. A. (1993) Using ARC/INFO and the National Wetland Inventory to model swamp rabbit (*Sylviilagus aquaticus) h*abitats, in *Proceedings of the 13th Annual ESRI Conference*, Redlands: Environmental Systems Research Institute, vol 1: 215 - 276.

Griffiths G.H., Smith, J.M., Vietch, N., and Aspinall, R. (1993) The ecological interpretation of satellite imagery with specific reference to bird habitats, in *Landscape Ecology and GIS*, ed. by R. Haines-Young, D.R. Green and S. Cousins, London: Taylor and Francis, pp. 225-271.

Hastie, T.J. (1993) Generalized additive models, in *Statistical Models in S*, ed. by J.M. Chambers and T.J. Hastie, London: Chapman and Hall, pp. 249-307.

Hastie, T.J., and Pregibon, D. (1993) Generalized linear models, in *Statistical Models in S*, ed. by J.M. Chambers and T.J. Hastie, London: Chapman and Hall, pp. 195-247.

Heliovaara, K., Vaisanen, R., and Immonen, A. (1991) Quantitative biogeography of the bark beetles (Coleoptera, Scolytidae) in northern Europe, *Acta Forestalia Fennica* 219: 1-35.

Hengeveld, R. (1994) Biogeographical ecology, *Journal of Biogeography* 21: 341-351.

Hill, M.O. (1991) Patterns of species distribution in Britain elucidated by canonical correspondence analysis, *Journal of Biogeography* 18: 247-255.

Hollander, A.D., Davis, F.W., and Stoms, D.M. (1994) Hierarchical representations of species distributions using maps, images and sighting data, in *Mapping the Diversity of Nature*, ed. by R.I. Miller, London: Chapman and Hall, pp. 71-88.

Hunsaker, C.T., Nisbet, R.A., Lam, D.C.L., Browder, J.A., Baker, W.L., Turner M.G., and Botkin, D.B. (1993) Spatial models of ecological systems and processes: The role of GIS, in *Environmental Modelling with GIS*, ed. by M.F. Goodchild, B.O. Parks and L.T. Steyaert, New York: Oxford University, pp. 248-264.

Johnston, C.A. (1993) Introduction to quantitative methods and modeling in community, population, and landscape ecology, in *Environmental Modelling with GIS*, ed. by M.F. Goodchild, B.O. Parks and L.T. Steyaert, New York: Oxford University, pp. 276-283.

Le Duc, M.G., Hill, M.O., and Sparks, T.H. (1992) A method for predicting the probability of species occurrence using data from systematic surveys, *Watsonia* 19: 97-105.

Lescourret, F., and Genard, M. (1994) Habitat, landscape and bird composition in mountain forest fragments, *Journal of Environmental Management* 40: 317-328.

Liebhold, A.M., Rossi, R.E., and Kemp, W.P. (1993) Geostatistics and geographic information systems in applied insect ecology, *Annual Review of Entomology* 38: 303-327.

Linzey, A., and Harvey, L.E. (1995) Modelling bird species distributions for gap analysis of the Tutamoe Ecological District, Northland, *Proceedings 7th AURISA/SIRC Colloquium*, Spatial Information Research Centre: Otago, pp.189-202.

Martin, J.-L., and Lepart, J. (1989) Impoverishment in the bird community of a Finnish archipelago: the role of island size, isolation and vegetation structure, Journal of Biogeography 16: 159-172.

Maurer B.A. (1994) *Geographical Population Analysis: Tools for the Analysis of Biodiversity*, Oxford: Blackwell.

Michaelsen, J., Schimel, D.S., Friedl, M.A., Davis, F.W, and Dubayah, R.C. (1994) Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys, *Journal of Vegetation Science* 5: 673-686.

Miller, R. I. (1994) Setting the scene, in *Mapping the Diversity of Nature*, London: Chapman and Hall, pp. 3-17.

Morrison, M.L., Marcot, B.G., and Mannan, R.W. (1992) *Wildlife-Habitat Relationships*, Madison: University of Wisconsin.

Ogle, C.C. (1981) The ranking of wildlife habitats, *New Zealand Journal of Ecology* 4: 115-123.

Okland, R.H., and Eilertsen, O. (1994) Canonical correspondence analysis with variation partitioning: some comments and an application, *Journal of Vegetation Science* 5: 117-126.

Osborne, P.E., and Tigar, B.J. (1992) Interpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa, *Journal of Applied Ecology* 29: 55-62.

Owen, J. (1990) An analysis of the spatial structure of mammalian distribution patterns in Texas, *Ecology* 71: 1823-1832.

Palmer, M.W. (1993) Putting things in even better order: the advantages of canonical correspondence analysis, *Ecology* 74: 2215-2230.

Pereira, J.M.C., and Itami, R.M. (1991) GIS-based habitat modeling using logistic multiple regression: A study of the Mt. Graham red squirrel, *Photogrammetric Engineering and Remote Sensing* 57: 1475-1486.

Porter, W.F., and Church, K.E. (1987) Effects of environmental pattern on habitat preference analysis, *Journal of Wildlife Management* 51: 681-685.

Poulin, B., Lefebvre, G., and McNeil, R. (1993) Variations in bird abundance in tropical arid and semi-arid habitats, *Ibis* 135: 432-441.

Rappole, J.H., Powell, G.V.N,. and Sader, S.A. (1994) Remote-sensing assessment of tropical habitat availability for a nearctic migrant: The wood thrush, in *Mapping the Diversity of Nature,* London: Chapman and Hall, pp. 91-103.

Rosen, B.R. (1988) Biogeographic patterns: a perceptual overview, in *Analytical Biogeography,* ed. by A. Myers and P. Giller, London: Chapman and Hall, pp. 23-55.

Sanderson, R.A., Rushton, S.P., Cherrill, A.J., and Byrne, J.P. (1995) Soil, vegetation and space: an analysis of their effects on the invertebrate communities of a moorland in north-east England, *Journal of Applied Ecology* 32: 506-518.

Smith, P.A. (1994) Autocorrelation in logistic regression modelling of species' distributions, *Global Ecology and Biogeography Letters* 4: 47-61.

Spitzer, K., Novotny, V., Tonner, M., and Leps, J. (1993) Habitat preferences, distribution and seasonality of the butterflies (Lepidoptera, Papilionoidea) in a montane tropical rain forest, Vietnam, *Journal of Biogeography* 20: 109-121.

Stoms, D.M. (1992) Effects of habitat map generalization in biodiversity assessment, *Photogrammetric Engineering and Remote Sensing* 58: 1587-1591.

Stoms, D.M. (1994) Scale dependence of species richness maps, *Professional Geographer* 46: 346-358.

ter Braak, C.J.F. (1987) *CANOCO - a FORTRAN program for canonical community ordination by (partial) (detrended) (canonical) correspondence analysis, principal components analysis and redundancy analysis (v2.1),* Netherlands: Agriculture Mathematics Group, Wageningen.

ter Braak, C.J.F. (1990) *Update notes: CANOCO version 3.10.* Netherlands: Agriculture Mathematics Group, Wageningen.

ter Braak, C. (1995) Ordination, in *Data Analysis in Community and Landscape Ecology*, ed. by R. Jongman, C. ter Braak and O. van Tongeren, Netherlands: Pudoc Wageningen, pp 91-173.

ter Braak, C., and Looman, C.W.N. (1995) Regression, in *Data Analysis in Community and Landscape Ecology,* ed. by R. Jongman, C. ter Braak and O. van Tongeren, Netherlands: Pudoc Wageningen, pp 29-77.

ter Braak, C.J.F., and Prentice, C. I. (1988) A theory of gradient analysis, in *Advances in Ecological Research* 18 : 271-317.

Thomas, D.L., and Taylor, E.J. (1990) Study designs and tests for comparing resource use and availability, *Journal of Wildlife Management* 54: 322-330.

Venables, W.N., and Ripley, B.D. (1994) *Modern Applied Statistics with S-Plus*, New York: Springer Verlag.

Vetaas, O.R. (1993) Effect of spatial arrangement of environmental variables on ordination results from a disturbed humidity gradient in northeastern Sudan, *Coenoses* 8: 27-37.

Walsh, S.J., and Davis, F.W. (1994) Applications of remote sensing and geographic information systems in vegetation science: introduction, *Journal of Vegetation Science* 5: 610 - 613.

Westman, W.E. (1991) Measuring realized niche spaces: climatic response of chaparral and coastal sage scrub, *Ecology* 72: 1678-1684.

Woodley, L.S., Francis, G., and Kay, J., eds. (1993) *Ecological Integrity and the Management of Ecosystems*, St. Lucie Press.

L. Edward Harvey.
Department of Geography, University of Auckland,
Private Bag 92019, Auckland, New Zealand.
ph: 64 9 373 7599 ext 8455
FAX: 64 9 373 7434
Email: e.harvey@auckland.ac.nz

Measuring and modeling (bio)diversity: an approach based on geographic, taxonomic and environmental relations

Richard J Aspinall, Diane M Pearson and Julia A Miller

1. Macaulay Land Use Research Institute
   Craigiebuckler, Aberdeen, Scotland, AB9 2QJ
   e-mail: richard@bamboo.mluri.sari.ac.uk
   e-mail: julia@bamboo.mluri.sari.ac.uk

2. Department of Geography
   Australian National University, Canberra, ACT 0200
   e-mail: diane@dart.anu.edu.au

Diversity is a key element in ecological understanding and measurement of diversity is increasingly important in modelling interactions between human activity and the state of the wider environment. Environmental processes and natural and human-induced disturbance produce pattern, and cause change, in diversity of the worlds resources at a range of spatial and temporal scales. The basis for measurement and analysis of biodiversity is, therefore, to understand taxonomic and environmental relationships of species and habitat variability at a range of spatial and temporal scales. In this paper we develop a spatially-based definition and analysis of diversity that includes taxonomic and geographic properties of environmental and biological variation. The performances of several existing ecological measures of diversity are tested using geographic data describing environmental and biological variation for Scotland. The concept of biodiversity can be treated hierarchically, and, as the response of an organism to environmental heterogeneity is spatially and temporally variable, this approach considers three criteria that are of particular importance:

1. behaviour as geographic scale changes,
2. behaviour in relation to hierarchical changes in taxonomic relations, and
3. the influence of spatial data quality

The pattern of diversity measured across Scotland is modelled using environmental data. This provides insights into the current environmental associations of ecological diversity at a range of geographic scales. This information can be particularly useful for models that predict the likely response of ecological processes to future disturbance. The use of diversity measures for environmental monitoring and assessment, and the values of diversity measures in spatial modelling are discussed.

Ecosystem Modeling of Spatially Explicit Land Surface Changes for Climate and Global Change Analysis

McKeown, R., Dennis S. Ojima, T.G.F. Kittel, D.S. Schimel, W.J. Parton, H. Fisher, T. Painter

Abstract

Ecosystem and atmospheric scientists are concerned with understanding how biospheric characteristics of the land surface will be modified in response to changing climate and land use. Ecosystem properties such as vegetation structure, carbon (C) fluxes, water exchange, and nitrogen (N) feedbacks to ecosystem dynamics respond differently when perturbed by climate or land use changes. These differences are reflected in both spatial and temporal features of the land surface characterization. Using ecosystem models to simulate changes in ecosystem properties due to land use and climate change perturbations across different land cover classes provides a way to test the sensitivity of different ecosystems.

Our analysis uses the spatially explicit version of the CENTURY ecosystem model to simulate ecosystem dynamics within a region. Site information including monthly rainfall, temperature, N deposition, soil properties, vegetation type and land-use management were defined for each grid cell. Simulation results based on a coarse grid regional representation vary significantly from those based on a finer grid. Attempting to resolve these discrepancies with simple spatial averaging of the driving variables often leads to conditions not actually observed in the region of interest. This led us to develop techniques to statistically represent sub-grid variability to better parameterize CENTURY and to assess changes in ecosystem dynamics resulting from changes in environmental factors. The results indicate that spatially-weighted averaging of results based on combinations of ecosystem drivers and properties which actually occur provide a better fit to observed ecosystem dynamics.

Introduction

The dominant factors controlling ecosystem dynamics are climate, nutrient availability, and land use. Our ability to predict changes in ecosystem dynamics and human welfare relative to climate or land use changes is dependent on the development of analytical tools to integrate our current understanding of how these ecosystems behave relative to human and environmental factors. The analysis of this information will need to incorporate critical factors of the physical environment (climate, land cover and soil) as well as factors defining human interactions with the environment (economic, social and cultural). The importance of past and current climate and land use cannot be overlooked in assessing how these ecosystems have developed over time and how they may change in the future relative to new policies, technological advances, economic conditions, and environmental constraints.

We used a spatially explicit version of the CENTURY model to examine the combined effects of climate and land use change on net carbon exchange in the central region of the US. Here, we present results of simulations performed at 0.5-degree resolution using soils and vegetation datasets that include a statistical representation of finer resolution information. Results of simulations made with and without sub-grid heterogeneity and with potential versus current vegetation cover are compared.

Model Description

The CENTURY ecosystem model Version 4 [Parton et al., 1987, 1993; Metherell, 1992] is a general model of plant-soil nutrient cycling which has been used to perform simulations across systems (including grasslands, agricultural lands, forests, and savannas) in various geographic regions. CENTURY is composed of several submodels: a soil organic matter / decomposition submodel, a water budget model, a grassland / crop production submodel, a forest production submodel, and management and events scheduling functions. The model computes the flow of C and N through a system of compartments using a monthly timestep. The following variables are required as input.

- Monthly average maximum and minimum air temperature.
- Monthly precipitation.
- Soil texture (% sand, silt, and clay).
- N content of plant material.
- Lignin content of plant material.
- Atmospheric and soil nitrogen inputs.
- Biome Type.

CENTURY incorporates simplified representations of key processes relating to carbon assimilation and turnover, including the impact of cropping, tillage, harvest, fire, grazing, and storm disturbances on ecosystems [Ojima et al., 1990; Sanford et al., 1991; Holland et al., 1992; Metherell, 1992]. Therefore CENTURY also requires parameters which specify disturbance types and frequencies for each biome.

As mentioned above, CENTURY consists of linked submodels. The soil organic matter (SOM) submodel simulates the decomposition of plant residues by microbes. The resulting microbial products become the substrates for SOM formation. The SOM is divided into three fractions: an active soil fraction representing live microbes and microbial products (1 to 4 yr. turnover time); a protected fraction representing the organic matter which is more resistant to decomposition as a result of physical or chemical protection (20 to 40 yr. turnover time); and a fraction that is physically protected or chemically resistant and has a long turnover time (800 to 1200 yr.). The water budget model calculates monthly evaporation, transpiration, the water content of the soil layers and snow, and saturated flow of water between soil layers. The plant production submodels both assume that the monthly maximum plant production is controlled by moisture and temperature, and that maximum plant production rates decrease if there are insufficient nutrient supplies. The grassland / crop production submodel simulates plant production for herbaceous crops and plant communities. The forest submodel simulates the growth of deciduous or evergreen forest in juvenile and mature phases. To simulate a savanna or shrubland, CENTURY uses both the grassland and forest submodels with some additional functions to perform nutrient competition and shading effects.

The recently developed spatially explicit version of CENTURY uses gridded maps of site specific driving variables as input (Figure 1). A simulation using this version of CENTURY begins by accessing the site information (climate, soil texture, and land use classification parameters) for a particular cell. Once the model has information on the land use classification for a cell, the model accesses the schedule file associated with that land use.

Information contained in the schedule file determines how CENTURY assigns management parameters, when events occur, and when the simulation ends.

CENTURY is an inherently transient model rather than an equilibrium model. In order to bring the C and N pools of simulated ecosystems to levels which are reasonable representations of existing ecosystems, CENTURY was run for at least 2000 years for each grid cell with prescribed disturbance regimes for specific biomes. Due to a lack of information about early site management, a generalized pattern for each system was modified so that the 2000 year run yields a representation similar to current conditions.

Geographical Data Bases

The domain for this model experiment was a region in the central grasslands of the United States and adjacent Rocky Mountains, including parts of Nebraska, Kansas, Colorado, and Wyoming. We made use of soil property (including soil texture and depth) and vegetation class data at two different scales of spatial resolution, 0.5-degree [from the VEMAP database, Kittel et al. 1995] and 10-km. The soil data were based on Kern's [1994, 1995] 10-km gridded Soil Conservation Service national-level (NATSGO) database. We used cluster analysis to group the 10-km sub-grid elements into 4 dominant soil types (modes) for each 0.5-degree cell. With this approach, soils properties for a 0.5-degree cell are represented by 1 or more dominant soil profiles rather than by an ``average soil profile" which may not correspond to an actual soil in the region.

The vegetation classes were those used by VEMAP [VEMAP Members 1995, Kittel et al. 1995]. The classes were defined physiognomically in terms of dominant life-form and leaf characteristics including leaf seasonal duration, shape, and size [Running et al. 1994]. In the case of grasslands, the classes were defined physiologically with respect to dominance of species with the C3- versus C4-photosynthetic pathway. Distribution of these classes was based on a 10-km gridded map of Kuchler's [1964, 1975] potential natural vegetation [NGDC in press]. For the purpose of this exercise, we assumed that this distribution of potential natural vegetation is in equilibrium with current climate. Current vegetation was estimated using the EROS Data Center 1-km land cover data for the US [Loveland et al. 1991], derived from AVHRR NDVI data. (Figure 2)

To avoid choosing vegetation-soil combinations which may not exist for a grid cell, actual vegetation-soil pairs were identified using bivariate histograms [Kittel et al. 1996, (Figure 3)]. The combination considered dominant was the most frequently occurring soil texture (%sand, %clay, bulk density) and the most frequently occurring land cover on that soil texture for each cell. The land cover and soil information was always coupled with the soil modes selected first and then 4 land cover modes selected for each soil mode. Simulations were run using the 16 most frequently occurring vegetation-soil combinations and the results combined on an area -weighted basis. This allowed us to include sub-grid information in our simulations without explicitly increasing the resolution of the simulations.

Methods

CENTURY simulations were run at a 0.5-degree resolution. Site input for a single grid cell's simulation consists of monthly precipitation, mean monthly maximum temperature, mean monthly minimum temperature, and a combination of land cover and soil texture. Each of the site input parameters is stored in a separate gridded map. The simulation of ecosystem dynamics for a single cell is independent of all other cells. To reduce I/O, CENTURY runs a single cell's entire scenario (all timesteps) before moving onto the next cell.

To account for sub-grid heterogeneity we ran 16 full grid simulations using identical climate information, each driven by different map sets containing existing land cover - soils combination. For all output variables, we used information on the area covered by the existing land cover - soils combinations to compute an area-weighted value for each cell.

Results: Land Cover Change

Across the region, potential natural vegetation based on the Kuchler vegetation classification was dominated by shortgrass steppe (roughly 60% of the area). Approximately 15% of the potential natural vegetation was classified as forest. The remaining area was occupied by cool-season grasses and tree-grass associations. The current land cover dataset indicates that cropland conversion resulted in a 33% transfer of the total land area into wheat-fallow / grassland mixture (class 118) mostly from shortgrass steppe.

Simulated soil C levels in the potential natural vegetation simulation and the simulation of current conditions were markedly different In the area converted to croplands from the shortgrass steppe, regional soil C levels averaged approximately 2.9 kg/m2. Higher soil C in croplands prior to conversion (~3.8 kg/m2) compared to shortgrass steppe not converted(~2.7 kg/m2) suggests that initial soil quality and organic matter content may have played a role in conversion management decisions (Figure 4).

Primary plant productivity for the region also indicated that differences between regions converted to cropland and non-converted areas existed. The Net Primary Production (NPP) for potential vegetation in grassland areas that were converted to croplands were about 72% greater than those areas not converted. This difference was greater than the relative change in NPP with cropping (Figure 5).

Results: Sub-grid Heterogeneity

When single dominant current vegetation and soil classes were simulated over the region, soil C levels declined by about 25% for the grid cells where cropland conversion took place. Soil C for the entire region dropped by about 10%. In simulations that included sub-grid information for land cover and soil, we saw an even greater loss in soil C. The grid cells in which land cover conversions took place showed a 32% loss in soil C from potential natural vegetation. An overall loss of 22% was seen for the entire region.

Conclusions

Land cover (vegetation), whether natural or human induced, is strongly correlated to environmental factors such as soil texture. In compiling data used to drive simulations these factors should not be decoupled. The non-linear behavior of ecosystem dynamics with respect to land cover change would lead to uncertain results. Using bivariate dataset development techniques, we were able to simulate sub-grid heterogeneity and maintain dependencies inherent in the data without increasing the resolution of the simulations. Although the technique we used required multiple 0.5-degree simulations, it still required fewer simulations than running at the resolution which was represented statistically by the bivariate data. Additional analysis of factors which determine the selection of areas for particular uses that go beyond the biophysical determinants also need to be incorporated into the data. This information is more difficult to quantify. However, efforts are underway to evaluate the contextual nature of social and economic factors in selection of areas for different land uses.

Acknowledgments

Support for this research was provided by the National Aeronautics and Space Administration's EOS Program through an EOS Interdisciplinary Science grant to David S. Schimel, and the VEMAP sponsors (the Electric Power Research Institute, the National Aeronautics and Space Administration, and the U.S. Department of Agriculture Forest Service). We also acknowledge the National Science Foundation support of the Climate System Modeling Program of the University Corporation for Atmospheric Research. We thank Nan Rosenbloom for her assistance in preparing datasets and post-processing.

References

Holland,E.A., W. J. Parton, J. K. Detling, and D. L. Coppock, Physiological response of plant population to herbivory and their consequences for ecosystem nutrient flow, Am. Nat., 140(4), 685-706, 1992.

Kern, J. S., Spatial patterns of soil organic carbon in the contiguous United States, Soil Sci. Soc. Am. J., 58, 439-455, 1994.

Kern, J. S., Geographic patterns of Soil water-holding capacity in the contiguous United States, Soil Sci Soc. Am. J., 59, 1126-1133, 1995.

Kittel, T. G. F., N. A. Rosenbloom, T. H. Painter, D. S. Schimel, and VEMAP Modeling Participants, The VEMAP integrated database for modeling United States ecosystem/vegetation sensitivity to climate change. J. Biogeography, 22, in press, 1995a.

Kittel, T. G. F., D. S. Ojima, D. S. Schimel, R. McKeown, J.G. Bromberg, T.H. Painter, N. A. Rosenbloom, W. J. Parton, and F. Giorgi, Model-GIS integration and dataset development for assessing the vulnerability of terrestrial ecosystems to climate change, in Proceedings Volume, NCGIA Second International Conference on Integrating GIS and Environmental Modeling, GIS World Books, Ft. Collins, Colo., 1996.

Kuchler, A. W., Potential Natural Vegetation of the Conterminous United States, manual to accompany the map, Spec. Publ. 36, 143pp., Am. Geogr. Soc., New York, 1964.

Kuchler, A. W., Potential Natural Vegetation of the United States, 2nd ed., map 1:3, 168,000, Am. Geogr. Soc., New York, 1975.

Loveland, T. T., J. W. Merchant, D. O. Ohlen, and J. F. Brown, Development of a land-cover characteristics database for the conterminous U. S., Photo. Eng. Rem. Sensing 57:1454-1463, 1991.

Metherell, A. K., Simulation of soil organic matter dynamics and nutrient cycling in agroecosystems, Ph.D. Dissertation, Colorado State University, Fort Collins, 1992.

NGDC. NOAA/EPA Global Ecosystems Database Project, Global Ecosystems Database Version 1.0, Disk B. National Geophysical Data Center, Boulder, Colo., Digital data, (in press).

Ojima, D. S., W. J. Parton, D. S. Schimel and C. E. Owensby, Simulated impacts of annual burning on prairie ecosystems, in Fire in North American Tallgrass Prairies, edited by S. L. Collins and L. L. Wallace 175 pp., Univ. of Oklahoma Press, Norman, Okla. and London, England, 1990.

Ojima, D. S., D. S. Schimel, W. J. Parton, and C. E. Owensby, Short- and long-term effects of fire on N cycling in tallgrass prairie. Biogeochemistry, 24:67-84, 1994.

Ojima, D. S., K. A. Galvin, and B. L. Turner II, The global impact of land-use change. BioScience, 44:300-304, 1994.

Parton, W. J., D. S. Schimel, C. V. Cole, and D. C. Ojima, Analysis of factors controlling soil organic matter levels in Great Plains grasslands, SSSAJ 51(5):1173-1179, 1987.

Parton, W. J., C. V. Cole, J. S. B. Stewart, D. S. Ojima, and D. S. Schimel, Simulating regional patterns of soil C, N, and P dynamics in the U.S. Central grassland region, in Ecology of Arable Land, edited by M. Clarholm and L. Bergstrom, pp. 99-108, Kluwer Academic Publishers, 1989.

Parton, W. J., et al., Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide, Global Biogeochem. Cycles, 7(4), 785-809, 1993.

Riebsame, W. E., W. J. Parton, K. A. Galvin, I. C. Burke, L. Bohren, R. Young, and E. Knop, Integrated modeling of land use and cover change. A conceptual scheme for applying and integration strategy to agricultural land use on the U.S. Great Plains, BioScience 44(5):350-356, 1994.

Running, S. W., T. R. Loveland, and L. L. Pierce, A vegetation classification logic based on remote sensing for use in global biogeochemical models, Ambio, 23(1), 77-81, 1994.

Sanford, Jr., R. L., W. J. Parton, D. S. Ojima, and D. J. Lodge, Hurricane effect on soil organic matter dynamics and forest production in the Luquillo Experimental Forest, Puerto Rico: Results of simulation modeling, Biotropica, 23:364-372, 1991.

VEMAP Participants, Vegetation/Ecosystem modeling and analysis project(VEMAP): Comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to simulate change and CO2 doubling, Global Biogeochem. Cycle, 9, 407-438, 1995

Author Information

Rebecca McKeown

Associate Scientist: National Center for Atmospheric Research (NCAR)
Research Associate: Natural Resource Ecology Laboratory (NREL)
Colorado State University, Ft. Collins, CO 80523
Phone: (970) 491-1623
FAX: (970) 491-1965
beckym@NREL.colostate.edu

Dennis Ojima
Research Scientist: NREL
Colorado State University, Ft. Collins, CO 80523
Phone: (970) 491-1976
FAX: (970) 491-1965
dennis@NREL.colostate.edu

Timothy Kittel
Deputy Project Scientist: Climate System Modeling Program (CSMP
Research Associate: NREL
University Corporation for Atmospheric Research
P.O. Box 3000, Boulder, CO 80307
Phone: (303) 497-1611
FAX: (303) 497-1695
kittel@ncar.ucar.edu

Thomas Painter
Graduate Student: Dept. of Geography
University of California, Santa Barbara
painter@crseo.ucsb.edu

Hank Fisher Programmer: CSMP
University Corporation for Atmospheric Research
P.O. Box 3000, Boulder, CO 80307
Phone: (303) 497-1617
FAX: (303) 497-1695
fisherh@ncar.ucar.edu

David Schimel
Scientist, Section Head: Climate and Global Dynamics Division (CGD)
Senior Scientist: NREL
University Corporation for Atmospheric Research
P.O. Box 3000, Boulder, CO 80307
Phone: (303) 497-1610
FAX: (303) 497-1695
schimel@ncar.ucar.edu

William Parton
Senior Scientist: NREL
Colorado State University, Ft. Collins, CO 80523
Phone: (970) 491-1987

FAX: (970) 491-1965
billp@NREL.colostate.edu

Last Modified: 04:29pm MDT, April 19, 1996

Miles G. Logsdon

Modeling Land-Cover Change From Measures of Spatial Landscape Structure

Abstract

Landscape level indices of spatial structure are used to examine forest and agricultural land cover transitions in time series data under conditions of varying spatial scale. Working with four coterminous classified TM images (August 1988, '89, '90 and '91) of an altered landscape north of the city of Manaus in northern Brazil, conditional probability matrices of land cover transition are compared to measures of landscape structure. A commercial GIS database and algorithms for modeling structural composition are used to define landscape elements and provide an estimate of the probabilistic behavior of pixels. While no strong relationship is found in the data presented, the conceptual framework and the GIS models developed in this investigation link probability estimates of pixel change with spatial measures of patch structure. This approach suggests a means of linking the probabilistic behavior of fine scale dynamics to the pattern observed at larger spatial scales.

I. Introduction

A landscape is composed of ever-changing elements. Their spatial and temporal patterns distinguish a landscape to an observer; at the same time they inform us of the complexity of dynamic processes at various scales. The changing pattern of the landscape, including the changing biophysical properties of that landscape, are a central theme in the fields of landscape ecology and environmental planning. Many of our research questions and management issues are focused upon the relationship between the changes that occur in the composition of the landscape and the spatial configuration of landscape elements.

The deforestation and fragmentation of tropical forests in the Amazon Basin of South America have become the focus of a global debate on international environmental policy (Porter and Brown 1991). The NASA Mission to Planet Earth - Earth Observing System (MTPE-EOS) research program provides the opportunity to investigate the hydrology, biogeochemistry and forest dynamics of the Amazon with multi-scaled remote sensing data (Richey et al. 1990). As an element of that program, this investigation explores a probabilisitic spatial model of landscape transition from measures of landscape structure where remotely sensed data and geographic information systems are integrated. This investigation explores the link between landscape patch structure and the individual pixel transition of intrapatch heterogeneity. Specifically, this investigation develops tools and tests the hypothesis that a probabilistic statement for the transition of a pixel, once the initial state of the pixel in a land cover time series dataset is known, can suggest the spatial structure of the landscape patch to which that pixel belongs.

It may help to illustrate this hypothesis with the following example. To an observer, a small cluster of contiguous cells classified as open canopy forest surrounded by a much larger and relatively homogeneous patch of primary forest may be recognized as a single landscape element of primary forest having some degree of intrapatch variety and a complex spatial structure. This investigation explores the hypothesis that the pixels of open canopy constitute pixels of fine scale intrapatch heterogeneity and that their probabilistic behavior for land cover change will be influenced by the compositional variety and spatial configuration of the larger scale pattern of the forest patch. Standard GIS functions are used to construct a four year time series spatial database of patches, probabilities of land cover transitions, and measures of spatial structure. This approach is discussed as a tool for modeling large scale spatial patterns from knowledge of finer scale dynamics.

II. Background

Simulating the stochastic nature of change has been of fundamental importance in ecology (Bell 1974, Pastor and Johnston 1992). As new perspectives in land management have emphasized planning of ecosystem sustainability (Franklin 1992), ecologists have begun to re-emphasize the role of spatial and temporal dynamics in their models (Baker 1989). As a result, spatially explicit stochastic simulation models have been applied to various landscapes and biophysical processes (Moloney et al. 1991, Turner 1987, Flamm and Turner 1994, Muller and Middleton 1994).

Multi-scale research has likewise received considerable attention of late. The non-cartographic meaning of the term "spatial scale" refers to geographic extent (window size) and resolution (the degree to which spatial objects are distinguishable). Scale then becomes a complex variable which captures the dynamics of change in both space and time. Theories and conceptual models of ecological scale (Allen and Star 1982, Meentemeyer and Box 1987, Lord and Norton 1990) have led to investigations of the simulation of multi-scale research (Delcourt and Delcourt 1988, Rastetter et al. 1992, Wessman 1992, Perestrello de Vasconcelos et al. 1993, King et al. 1991). Others have gone on to look at the application of various methods for modeling multi-scale data (Gaines & Denny 1993, Smith & Urban 1988, Carlile et al. 1989, Turner and Gardner 1991, O'Neill et al. 1991).

The essential goal of modeling and monitoring environmental change from remotely sensed data is to compare images at a spatial and temporal resolution appropriate to the ecological scale of the processes of interest. Satellite remote sensing instruments provide measurements at a variety of pixel resolutions, spatial extents and temporial scales. However, due to variability in illumination, atmospheric effects, and instrument calibration, conventional supervised or unsupervised classification techniques have difficulty providing pixel to pixel comparisons between images from different times. Classification of any given pixel into a discrete land cover class for the purpose of determining change requires that these variables be considered (Adams et al. in press).

After arriving at comparable classes, current work in change detection has begun to go beyond simple descriptive summations of change. LaGro and DeGloria (1992) applied multiple regression to modeling land use dynamics where the proportions of change in each of five classes of land use/land cover were used as the dependent variable. The error term for each linear regression was attributed to sampling errors, variables not used in the analysis, or stochastic effects. While land cover change has received less attention in spatially explicit models than other biophysical processes, this and other works have suggested that current models are insufficient for the complexity of anthropogenic impacts. Where anthropogenic forces are at work land cover change should be thought of in probabilistic terms.

Conceptual framework

We can arrive at a probabilistic statement concerning the transition of any pixel with a conceptual model of the interaction between the current land cover state of the pixel, the spatial context (patch membership) of the pixel, and the likelihood of transition. The transition of a pixel from one discrete land cover class to another is a function of the present land cover class of that pixel, the structural composition and configuration of the patch to which that pixel belongs, and the conditional probability of pixels of that class making transitions between classes. The structural configuration and composition of the patch to which the pixel belongs is a function of the shape of the patch (configuration) and the variety of unique land cover types (composition) found within the patch (Dunning et al. 1992, Turner 1989). The probabilistic term is captured in a matrix of conditional probabilities.

This general concept can be expressed clearly in two statements. The first recognizes the influence of patch stucture on the future behavior of individual member pixels along with the influence of all other non-defined processes (E), while the second recognizes that the transition is governed by random variables and may be described only in probabilistic terms.

$$LC^{m}_{t+1} = f(C[patch^{m}_{i,k}]_{t}, E) \qquad (2.1)$$

and also,

$$P(LC^{m}_{t+1} = j \mid LC^{m}_{t} = i, LC^{m}_{t-1} = i_{1}, ..., LC^{m}_{0} = i_{0})$$

$$= P\{LC^{m}_{t+1} = j \mid LC^{m}_{t} = i\} = P_{i,j} \qquad (2.2)$$

Where:

m = 1 ..., n cells (pixels)
i = 1 ..., 8 patch type (classes)
j = 1 ..., 8 patch type (classes)
k = 1 ..., *l* patches
t = t, 1, 2, ... time

And:

LC^{m}_{t} is the land cover class of pixel m at time t (2.3)

and:

m

$$LC_t = 1 \ldots, 8 \text{ for all } m$$

And:

(2.4) $C[patch^m_{i,k}]_t$ is the structure of the kth patch of class i where pixel m is a member at time t

Where:

$$C = f(D, VAR)$$

Such that:

$$D = \frac{2 \ln(.25 \text{ perimeter})}{\ln \text{area}}$$

and:

VAR is the unique number of i's in the patch

Equation 2.2 expresses the Markov assumption that the future behavior of a pixel is determined once the state of the process at the present time is known. $P_{i,j}$ stand for the conditional probability for change of a given pixel in state i at time t to state j at time t+1. $P_{i,j}$ is derived from the initial change matrix by simply totalling the rows of the joint probabilities and dividing each element in the row by its own row total. When the probabilities have been normalized, the conditional probabilities form the transition, or stochastic, matrix. The transition mechanism of individual pixels is captured in the matrix of conditional probabilities.

The two earlier conceptual statements (2.1 and 2.2) express patch structure (2.4) as a strong determinant in the future state of a given pixel. This investigation tests the general hypothesis that patch structure can be used to model landcover transitions. More specifically it tests the hypothesis that the probability of transition will increase as patch structure becomes more complex. Or,

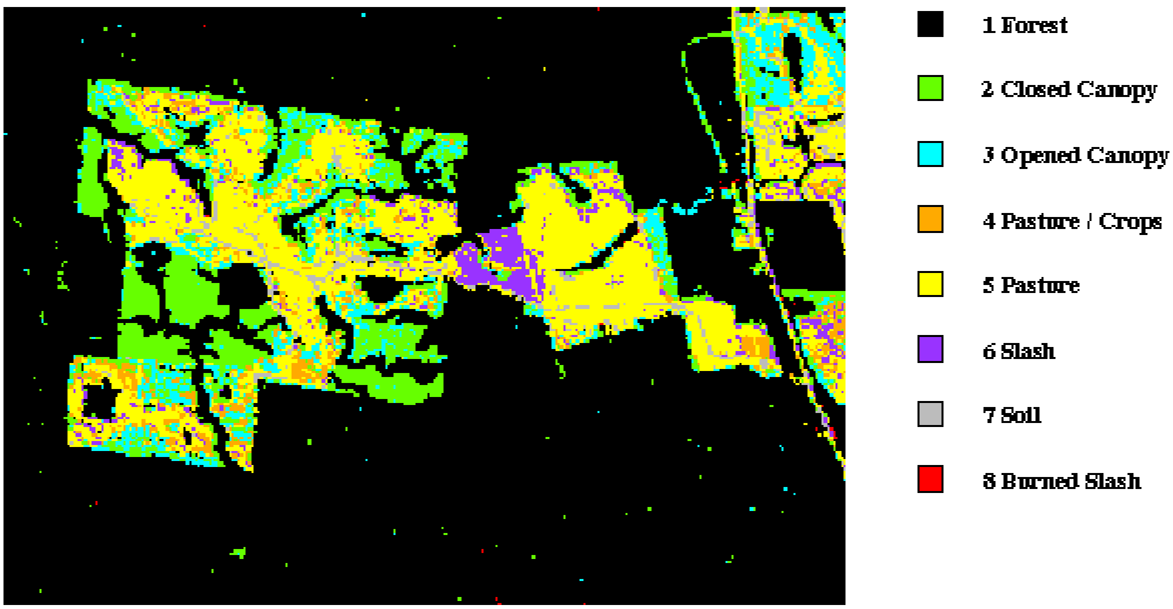
(2.6) as $C[patch^m_{i,k\,t}]$ increases, then $P_{i,j} \rightarrow 1$

III. Methods

The commercial geographic information system (GIS) software Arc/Info ver 7.0.1 produced by Environmental Systems Research Institute (ESRI) was used in this investigation. The spatial models rely heavily upon the use of raster based spatial model language, an implementation of the map algebra functions and the Arc Marco Language (aml) of Arc/Info. A minimum of specialized programming code and operating system scripts were used when the macro language was inadequate or cumbersome.

The investigation begins with four spatially coterminous land cover classified Landsat Thematic Mapper (TM) images (August 1988, '89, '90 and '91) of the Basin of Greater Manaus in northern Brazil (3° South, 60°West). The classified land cover dataset (Figure 1) is provided by the NASA supported EOS-Amazon project team and derived outside the scope of this investigation (see Adams et al. in press). Each pixel was assigned to a land cover class through application of a spectral mixing model. This procedure provides the fractional composition (percent of each pixel) for one of four end-members (non-photosynthetic vegetation, photosynthetic vegetation, soil and shade). The classification step concludes by assigning one of eight recognizable land cover "labels" to each pixel.

Figure 1
Land Cover Classification

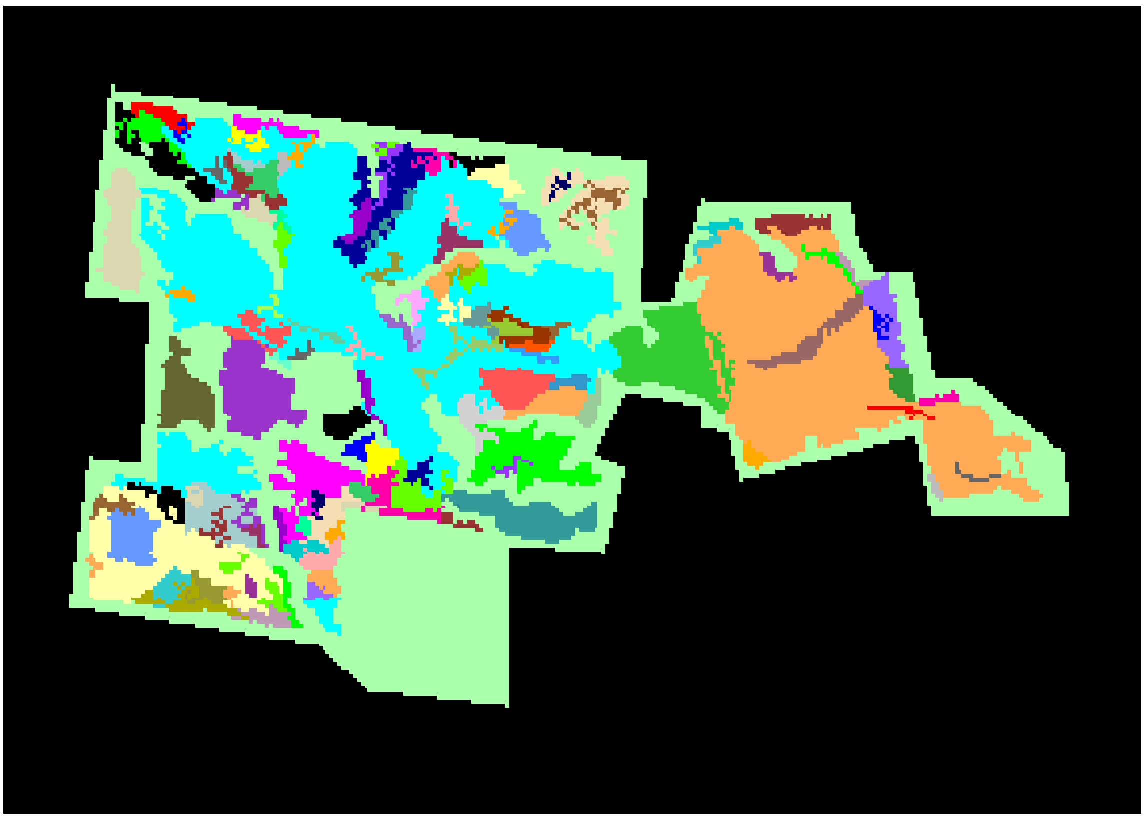


The GIS dataset is constructed by clipping these images to a common spatial extent and converting them to raster GIS data model (grids) with a cell size of 30m X 30m. The study area is in a region of the Amazon basin where forest clearing for agriculture began in the late 1970's. This area, referred to as Fazenda Dimona, began as a cattle ranch and now includes the research areas of the Biological Dynamics of Forest Fragments Project (BDFFP) of the National Institute for Amazon Research (INPA) and the Smithsonian Institution (Lovejoy and Bierregaard 1990). The management history of the site (Bierregaard et. al. 1992) has resulted in a variety of land covers from well-maintained pasture through regrowth to primary forest (Lucas et al. 1993).

The value of each cell in the land cover grids represents that cell's land cover value at the first time step (eq. 2.3). The spatial connectivity of these values and a variable for the allowable size of a landscape patch are used to construct landscapes of largely homogeneous patches with attributes of intrapatch heterogeneity. In this way the resolution, a component of spatial scale, is defined by the analyst. In this investigation the minimum patch size is one hectare.

Landscape patches are constructed through a step model which uses a number of map algebra steps to evaluate the size of discrete regiongroups (patches) and, based upon a user defined value for the allowable size of a minimum sized patch, creates patches which can contain intrapatch heterogeneity. In this way patches are recognized as landscape elements containing heterogenous land cover types. For example, a pasture may represent as a single landscape element but recognized as a patch containing small groupings of non-pasture land cover. The land cover value for individual pixels (eq. 2.3) of patches that fall below the allowable minimum patch size are replaced with the land cover value of the nearest landscape patch. New regiongroups are defined and a spatial union is performed to link the newly defined patches to their original patch identities. In this way only those cells which were in patches that fell below the minimum value will have different values of patch identity in this final step. Small patches and isolated cells are therefore assigned to the nearest landscape patch (Figure 2).

Figure 2
Landscape Patches (minimum allowable size of 1 hectare)



Various measures of spatial structure, including the value for D in equation 2.4, can be obtained simply from software applications such as Fragstats ver. 2.0 (McGarigal and Marks 1994). However, in this investigation values of both terms in equation 2.4 are generated by a macro using grid operators and the functions of zonalarea, zonalperimeter, and zonalvariety, where the zone was defined as a regiongroup and forms unique patches. A spatial union joins the grids on a cell by cell basis, while preserving in its attribute table the patch-ID and values for the terms D and VAR of each patch.

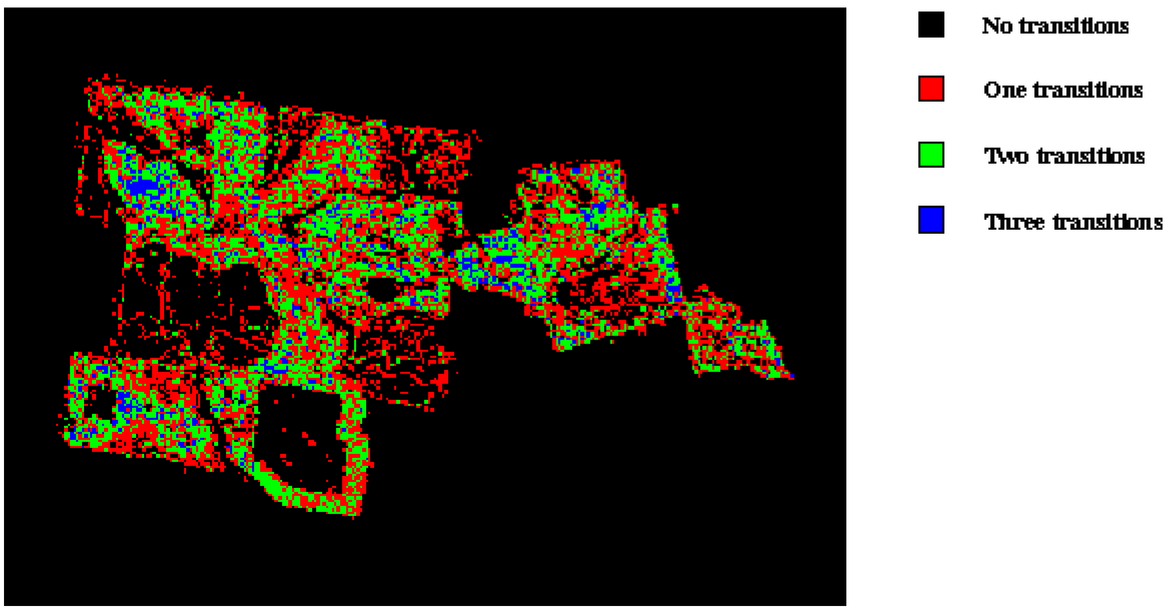
The initial change matrix is constructed from the attribute table of a grid resulting from a spatial union between each time step. A C program then creates the matrix of conditional probabilities (eq. 2.2) by totalling the rows of the joint probabilities and dividing each element in the row by its own row total.

The final operation is to join the database files for the grids containing the values of patch structure with the grid containing unique transitions within each patch, on the common item of patch-ID.

IV. Results

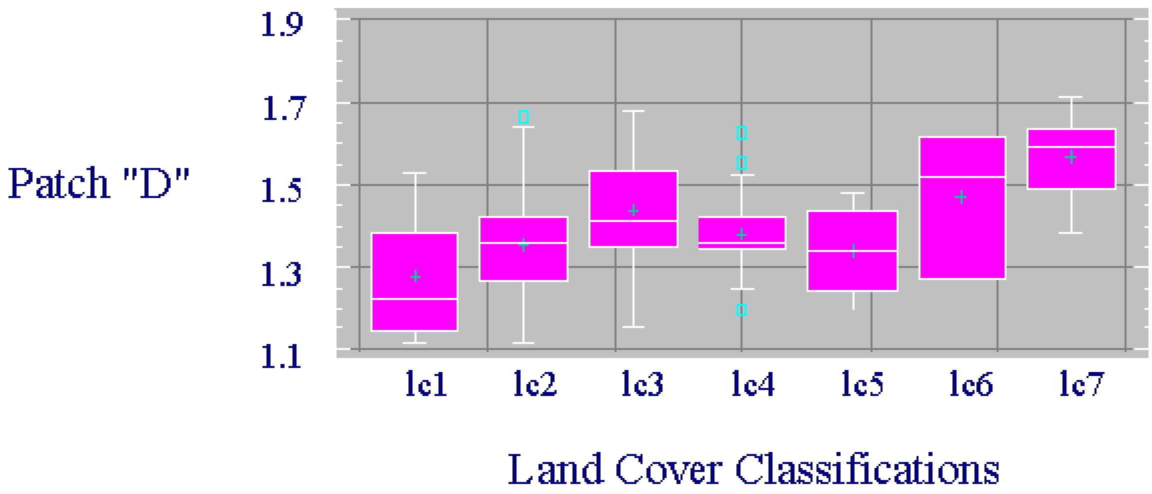
Temporal representation of variety among yearly grids (Figure 3) begins by assigning to each cell a value equal to the number of unique values for that cell in the set of yearly grids (i.e. a range of 1 - 4). This value describes the degree of variation in the transition for each cell in the time series. It does not inform us about the path or mechanism of the transition, but only provides a spatial representation of the variety found within each cell's temporal domain.

Figure 3
Land Cover Transitions (number of land cover values within the dataset)



The measure of patch structure (2.4) includes a term for patch configuration (D) described here as a simple 2-dimensional fractal measure in an area to perimeter ratio. Each patch is modeled as a homogeneous patch composed of one of eight land cover types. A value of D is unique to each patch. A value of 1.0 corresponds to simple shapes (circles or squares) and increases towards 2.0 as the shape becomes more highly convoluted (Figure 4).

Figure 4
Characteristics of D by Land Cover Types



The matrix of conditional probabilities is derived from the land cover change matrix. The matrices in Table 1 display the likelihood of transition between states based upon the relative frequency in one time step, when that time step was one year (1988 to 1989). It is characteristic of land cover transition matrices that the probabilities on the matrix diagonal (no change in state) are usually greater than o.5. This reflects that for many land cover types there is a relatively high likelihood that no change will occur. Table 1(a) reports the estimates of conditional probability based upon the behavior of all cells in the landscape, while Table 1(b) reflects that probability from only those cells which belonged to patches that were below the allowable patch size and were therefore modeled as cells of intrapatch heterogeneity.

Table 1
Conditional Probabilities Matrices ($P_{i,j}$)
(one year time step)

		1989							
		1	2	3	4	5	6	7	8
1988	1	0.913	0.056	0.013	0.002	0.008	0.004	0.001	0.003
	2	0.066	0.803	0.048	0.019	0.023	0.032	0	0.009
	3	0.084	0.479	0.225	0.041	0.085	0.064	0.002	0.020
	4	0.027	0.318	0.221	0.209	0.138	0.081	0.005	0
	5	0.022	0.045	0.092	0.115	0.577	0.118	0.030	0.002
	6	0.028	0.149	0.041	0.291	0.300	0.184	0.006	0.001
	7	0.024	0.002	0.022	0.022	0.467	0.009	0.453	0
	8	0.250	0.250	0	0.250	0	0	0	0.250

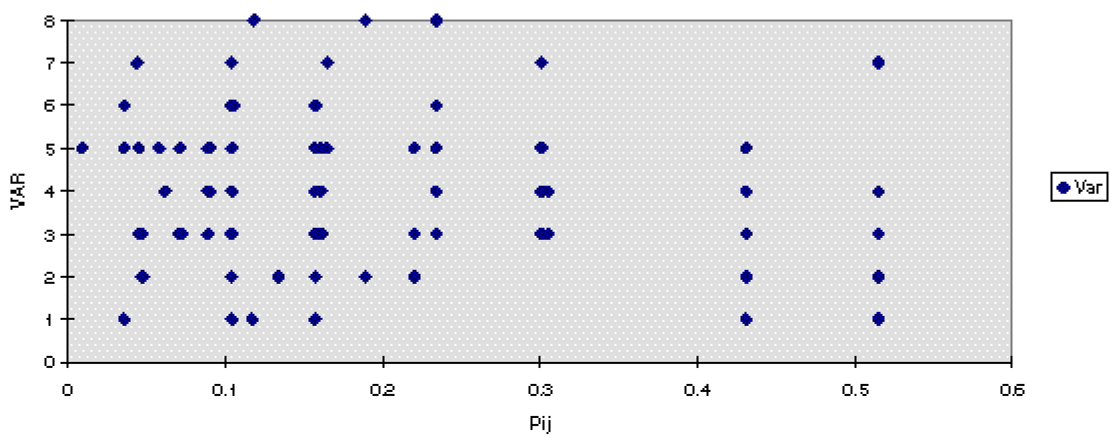
(a) Conditional probabilities derived from all landscape cells

		1989							
		1	2	3	4	5	6	7	8
1988	1	0.559	0.126	0.045	0.054	0.072	0.117	0.018	0.009
	2	0.157	0.525	0.104	0.071	0.062	0.058	0	0.023
	3	0.134	0.431	0.249	0.047	0.089	0.036	0.004	0.009
	4	0.034	0.301	0.234	0.159	0.161	0.105	0.004	0.001
	5	0.118	0.165	0.220	0.090	0.317	0.047	0.040	0.003
	6	0.044	0.189	0.048	0.305	0.215	0.189	0.008	0.002
	7	0.028	0.003	0.031	0.031	0.515	0.009	0.383	0
	8	0.250	0.250	0	0.250	0	0	0	0.250

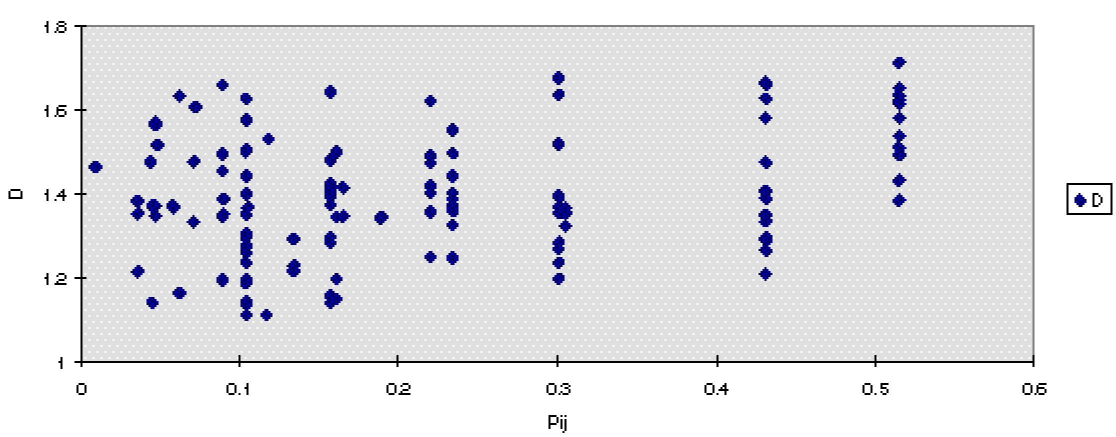
```
(b) Conditional probabilities derived from cells assigned as
    cells of intrapatch heterogeneity
```

The goal of this investigation was to explore the relationship between patch structure and the probability of land cover transition of individual cells ($P_{i,j}$). These probabilities are spatially linked to activity between cells in the GIS database and therefore to patch membership. Figures 5(a) and 5(b) graph the values of landscape patch composition (VAR) and configuration (D) against the conditional probability of making a transition during a one year time step, calculated from cells of intrapatch heterogeneity. In this way, the probability estimate reflects the finer spatial scale dynamics, while models of patch composition and configuration represent the large scale structure.

Figure 5
Comparison of Patch Structure and Probability of Cell Transition
(one year time step)



(a) Comparison of patch composition (the variety of intrapatch heterogeneity) and the probability of member cells making a transition.



(b) Comparison of patch configuration (a two-dimensional fractal measure) and the probability of member cells making a transition.

One important issue not addressed in this investigation is the nonstationary (non-temporally homogeneous) nature of many ecological Markov models. The ecological processes that are reflected by transitions may not be constant in space or time (Bell 1974). Therefore, we would expect the estimate of conditional probabilities calculated from observations of a three year time step (Table 2) to differ from those of a one year time step (Table 1) raised to the third power. In this investigation the nonstationary nature of the processes is ignored and the matrices are used as estimates of probability for discrete time steps. Table 2 and Figure 6 are the results of calculating conditional probabilities from a three year time step and comparisons to patch structure at time t.

Table 2
Conditional Probability Matrix ($P_{i,j}$)
(Three year time step)

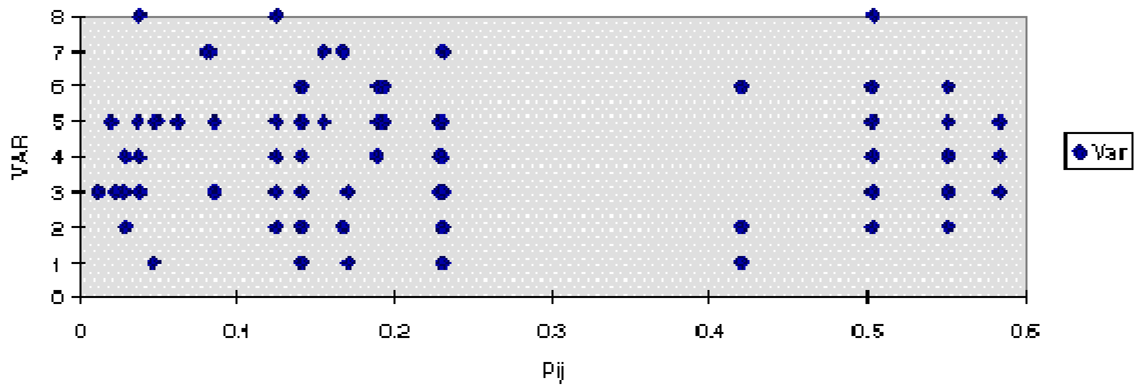
		1991							
		1	2	3	4	5	6	7	8
	1	0.840	0.082	0.012	0.011	0.011	0.039	0	0.004
	2	0.082	0.806	0.070	0.010	0.020	0.011	0	0
	3	0.147	0.639	0.155	0.017	0.037	0.006	0	0
1988	4	0.036	0.609	0.205	0.053	0.088	0.005	0.004	0
	5	0.040	0.301	0.240	0.085	0.248	0.073	0.012	0
	6	0.030	0.568	0.194	0.054	0.100	0.050	0.005	0
	7	0.041	0.119	0.205	0.039	0.421	0.026	0.149	0
	8	1.000	0	0	0	0	0	0	0

(a) Conditional probabilities derived from all landscape cells

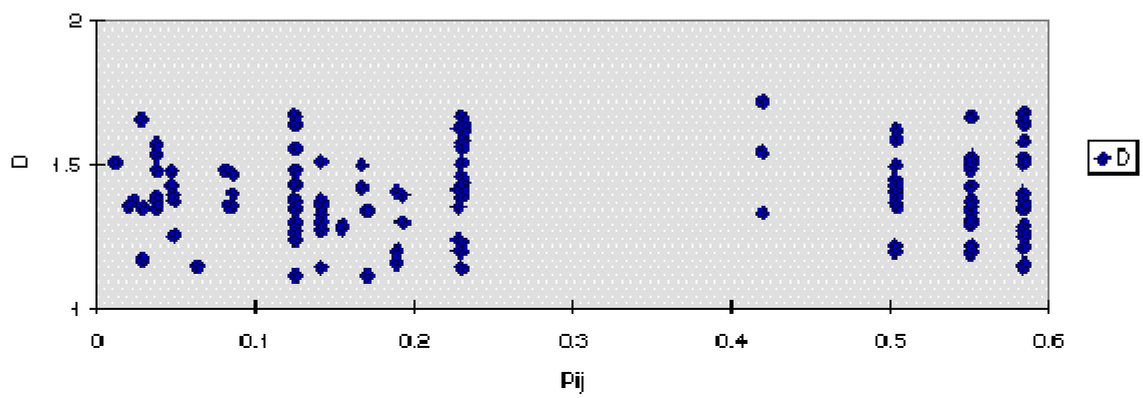
		1991							
		1	2	3	4	5	6	7	8
	1	0.622	0.171	0.081	0.018	0.063	0.045	0	0
	2	0.141	0.635	0.125	0.029	0.047	0.023	0	0
	3	0.230	0.551	0.152	0.020	0.038	0.009	0	0
1988	4	0.049	0.584	0.228	0.044	0.083	0.006	0.006	0
	5	0.193	0.503	0.189	0.012	0.065	0	0.037	0
	6	0.034	0.504	0.167	0.086	0.155	0.046	0.008	0
	7	0.037	0.154	0.231	0.025	0.420	0.028	0.105	0
	8	1.000	0	0	0	0	0	0	0

(b) Conditional probabilities derived from cells assigned as cells of intrapatch heterogeneity

Figure 6
Comparison of Patch Structure and Probability of Cell Transition
(Three year time step)



(a) Comparison of patch composition (the variety of intrapatch heterogeneity) and the probability of member cells making a transition.



(b) Comparison of patch configuration (a two-dimensional fractal measure) and the probability of member cells making a transition.

V. Conclusion

There appears to be no strong relationship between the measure of patch composition (VAR) or configuration (D) and the conditional probabilities of transition ($P_{i,j}$) of pixels modeled as pixels of intrapatch heterogeneity for either the one year or three year time step. The matrices of conditional probability reveal differences between the estimates of probabilistic behavior for pixels of intrapatch heterogeneity and all landscape pixels. These differences can be noted by comparing the probabilities on the matrix diagonal (no change in state) and provide a feel for the increased probability for change in pixels of intrapatch heterogeneity.

The conceptual framework presented and the GIS models developed in this investigation have linked probability estimates of pixel change with spatial measures of patch structure. This approach provides a means to link the probabilistic behavior of fine scale dynamics to the spatial pattern observed at larger spatial scales. Future work should explore the temporal homogeneous nature of the conditional probability matrix if the Markovian property is to be used as a predictive tool. Finally, different minimum allowable patch sizes and indices of patch structure may result in a stronger relationship between pixel transition and landscape structure.

Acknowledgments: *The digital classified images were provided by the staff of the Geological Remote Sensing Laboratory at the University of Washington. This investigation itself was made possible by support from the NASA MTPE-EOS Amazon Project. Technical assistance in programming and GIS functions was provided by Emilio Mayorga and Harvey Greenberg. General review and assistance in the preparation of this paper was provided by Dr. Earl J. Bell and Josh Greenberg, at the University of Washington.*.

References

Adams, J.B., D. Sabol, V. Kapos, R.A. Filho and others. In press. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon.*Remote Sensing of Environment*.

Allen, T.F.H. and T. Starr. 1982. *Hierarchy: Perspectives for Ecological Complexity*. The University of Chicago Press. Chicago.

Baker, W.L. 1989. A review of models of landscape change*Landscape Ecology* . 2: 111-113.

Bell, E.J. 1974. Markov analysis of land use change-- an application of stochastic processes to remotely sensed data. *Socio-Economic Planning Science* 8: 311-316.

Bierregaard, R.O.J., T.E. Lovejoy, V. Kapos, A.A.d. Santos and others. 1992. The biological dynamics of tropical rain forest fragments. *Bioscience* 42: 859-866.

Carlile, D.W., J.R. Skalski, J.E. Batker, J.M. Thomas and others. 1989. Determination of ecological scale *Landscape Ecology* 2: 203-213.

Delcourt, H.R. and P.A. Delcourt. 1988. Quaternary landscape ecology: Relevant scales in space and time. *Landscape Ecology* 2: 23-44.

Dunning, J.B., B.J. Danielson and H.R. Pullian. 1992. Processes that effect populations in complex landscapes. *Oikos* 65: 169-175.

Flamm, R.O. and M.G. Turner. 1994. Alternative model formulations for a stochastic simulation of landscape change. *Landscape Ecology* 9: 37-46.

Franklin, J.F. 1992. Scientific basis for new perspectives in forests and streams, p. 25 - 72. In: R.J. Naiman (ed.), *Watershed Management: Balancing Sustainability and Environmental Change*. Springer-Verlag. New York.

Gaines, S.D. and M.W. Denny. 1993. The largest, smallest, highest, lowest, longest, and shortest: Extremes in ecology. *Ecology* 74: 1677-1692.

King, A.W., A.R. Johnson and R.V. O'Neill. 1991. Transmutation and functional representation of heterogeneous landscapes. *Landscape Ecology* 5: 239-253.

LaGro, J.A. and S.D. DeGloria. 1992. Land use dynamics within an urbanizing non-metropolitan county in New York State (USA). *Landscape Ecology* 7: 275-289.

Lord, J.M. and D.A. Norton. 1990. Scale and the spatial concept of fragmentation. *Conservation Biology* 4: 197-202.

Lovejoy, T.E. and R.O.J. Bierregaard. 1990. Central Amazonian forests and the minimum critical size of ecosystems project. In: A. Gentry (ed.), *Four Neotropical Rain Forests*. Yale University Press. New Haven, CT.

Lucas, R.M., M. Honzak, G.M. Foody, P.J. Curran and others. 1993. Characterizing tropical secondary forests using multi-temporal Landsat sensor imagery. *Int. J. Remote Sensing* 14: 3016-3067.

McGarigal, K. and B.J. Marks. 1994. *Fragstats: Spatial Pattern Analysis program for Quantifying Landscape Structure*. Oregon State University. Corvallis, Oregon.

Meentemeyer, V. and E.O. Box. 1987. Scale Effects in landscape studies, p. 15-34. In: M.G. Turner (ed.), *Landscape Heterogeneity and Disturbance: Ecological Studies*. Springer-Verlag. New York.

Moloney, K.A., A. Morin and S.A. Levin. 1991. Interpreting ecological patterns generated through simple stochastic process. *Landscape Ecology* 5: 163-174.

Muller, M.R. and J. Middleton. 1994. A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. *Landscape Ecology* 9: 151-157.

O'Neill, R.V., S.J. Turner, V.I. Cullinan, D.P. Coffin and others. 1991. Multiple landscape scales: An intersite comparison. *Landscape Ecology* 5: 137-144.

Pastor, J. and C.A. Johnston. 1992. Using simulation models and geographic information systems to integrate ecosystem and landscape ecology, p. 324 - 346. In: R.J. Naiman (ed.), *Watershed Management: Balancing Sustainability and Environmental Change*. Springer-Verlag. New York.

Perestrello de Vasconcelos, M.J., B.P. Zeigler and L.A. Graham. 1993. Modeling multi-scale spatial ecological processes under the discrete event systems paradigm. *Landscape Ecology* 8: 273-286.

Porter, G., and J. W. Brown, 1991. *Globally Environmental Politics*. West View Press. Boulder, CO.

Rastetter, E.B., A.W. King, B.J. Cosby, G.M. Hornberger and others. 1992. Aggregating fine-scale ecological knowledge to model coarser-scale attributes of ecosystems. *Ecological Applications* 2: 55-70.

Richey, J.E., J.B. Adams and R.L. Victoria. 1990. Synoptic-scale hydrological and biogeochemical cycles in the Amazon river basin: A modeling and remote sensing perspective, p. 249 - 268. In: R.J. Hobbs and H.A. Mooney (ed.), *Remote Sensing of Biosphere Functioning*. Springer-Verlag. New York.

Smith, T.M. and D.L. Urban. 1988. Scale and resolution of forest structural pattern. *Vegetatio* 74: 143-150.

Turner, M.G. 1987. Spatial simulation of landscape changes in Georgia: A comparison of 3 transition models. *Landscape Ecology* 29-36.

Turner, M.G. 1989. Landscape ecology: the effect of pattern on process. *Annu. Rev. Ecol. Syst.* 20:171-197.

Turner, M.G. and R.H. Gardner (ed.). 1991. *Quantitative Methods in Landscape Ecology: The Analysis and Interpretation of Landscape Heterogeneity*. Springer-Verlag. New York.

Wessman, C.A. 1992. Spatial scales and global change: Bridging the gap from plots to GCM grid cells. *Annu. Rev. Ecolo. Syst.* 23: 175-200.

Miles G. Logsdon
EOS-Amazon Project
University of Washington
School of Oceanography
P.O. Box 357940
Seattle, WA 98195-7940
Telephone:(206)543-5334
e-mail: mlog@u.washington.edu

Thomas C. Hart, Stephanie L. Greene, Alexandr Afonin

Mapping for Germplasm Collections: Site Selection and Attribution

Abstract

Conservation of agricultural germplasm involves complex evaluation and maintainance of existing collections, in addition to acquisition through well-planned collection expeditions. This case study covers the use of modelled ecogeographic attributes to support a US/Russian expedition to collect forage legume germplasm in the western Caucasus Mountains during the summer of 1995. To isolate promising areas with a variety of habitats sampled within a general focus of tolerance to acid soils, the team combined examination of extensive and deductive habitat descriptions with more traditional hunting approaches.

Eighty map attributes were compiled on 500 meter cells, including longterm monthly moisture and temperature characteristics, terrain description, and soil character. Climate interpolations were kriged from over 100 stations distributed at an average interval of 30 kilometers, and were corrected for elevation effects via normalization and recalculation with an input DEM. Moisture and temperature zones were derived by ISODATA clustering on sets of nine and six seasonal combinations. There were mixed results with respect to the reliability of the data preparations.

More important was the progression of map data assimilation on the part of the field scientists. The modeled attributes provided a set of field corroborations which gave confidence and useful inference to the team members. Sampling gradients were devised with targets changing as the needed sets were filled. Experimental combinations were abandoned with improved confidence after repeated attempts to discover suitable sites. Nearly 600 seed samples were gathered with a better knowledge of their bias relative to natural distributions and abundances. Map query yielded attribute estimates that greatly extended each sample's passport database within the US and Russian seed archive systems.

Background

The Challenge of Sampling Agricultural Biodiversity

There is today a vitally important international effort to improve agricultural crops by breeding new varieties which are more tolerant to environmental stresses. Given global trends of reducing agricultural inputs and of farmland expansion into marginal areas, the search for germplasm adapted to suboptimal conditions is a growing priority. In the mid 1990s many crops remain limited by weaknesses in their winter hardiness, their drought tolerance, and their acid or salt tolerance. When breeders do achieve a substantial improvement on a common crop, there often are benefits measurable in millions of dollars, and in better nutrition for the impoverished. In the case of alfalfa, a century ago the crop was grown in only limited areas of the US due to low winter hardiness. Currently the widespread high yields of alfalfa are explained to a large extent by the introduction of winter-hardy germplasm from Northern Europe, Russia, Kashmir, and Turkistan (Bolton et al 1972).

Breeding programs make use of large archives of germplasm maintained by national and international genebanks. These institutions are responsible for the collecting, cataloguing, protection, and scientific access of literally millions of germplasm samples from tens of thousands of crop varieties. Unfortunately many national genebanks are under heavy budgetary pressure. This is especially true for the V.I. Vavilov Institute of Plant Industry in Russia, which houses one of the most significant and largest germplasm collections in the world (NRC 1993). Inasmuch as every nation shares a future where each collection may prove critical to future productivity, the collaboration between countries supports an enlightened common interest. For the past five years the USDA National Plant Germplasm System (NPGS) has been providing support to the Vavilov Institute.

Under this momentum during the summer of 1995 a major collecting trip was launched to acquire wild forage species (alfalfa, clover, grasses) in the Western Caucasus Mountains of Southern Russia. The mission's primary goal was to sample populations adapted to the acidic soil regions of the area. The objectives of two national programs and eight scientists from the disciplines of plant breeding, botany and soils, compounded by a field schedule of three months duration within the highly diverse environments of the Caucasus, resulted in an innumerable set of options from both biological and logistical perspectives. Careful planning was imperative to enable an efficient and focussed sampling strategy.

The traditional approach typically builds on known botanical geography. Trip preparation usually includes discussions with local authorities and examination of herbarium specimens to gain better understanding of the local environments and the predicted distributions of flora (Reid and Strickland 1983). The field effort then forms a series of hunting forays where collecting is guided by the progressive enroute discovery of plant associations, and by a rhythmic sampling to capture the decay of spatial autocorrelation. In the case of the Caucasus expedition, the ambitious breadth of science and geography led to an early commitment towards habitat-oriented maps to help guide the selection of target areas and sort through the long list of priorities. There was an initial desire to construct a set of all purpose strata, to be called Terrain Units, to serve as a sampling frame and a multi-attribute site classifier.

Application of Modeling and GIS

Historically the motivation for collecting expeditions was built on the perception of gaps in the existing collections, where summary taxonomic and geographic distributions had raised expectations that there was yet uncollected plant material which exhibited resistence to disease, insect and environmental stress. Ideally such a program would be designed to round out a species' representation by sampling throughout its diverse habitats (Brown and Marshall 1995). The strategy on this project was to search for such genes tolerant to acidic soils in regions (and on sites) which displayed strongly the unfavorable soils, with the hope that natural selection had induced a useful set of genotypes. However, a general weakness in habitat descriptions for most archived accessions has resulted in an incomplete picture of the degree to which each taxon's diversity is already represented within the collections. This is a situation that GIS and modeling can ameliorate: the portions of the collections with locator information can be attributed with habitat descriptors from the array of historic and recent map-based compilations. For some taxa this would provide a near-complete picture, but for others there are high proportions of geographically "lost" accessions, so that this mechanism would only hold minor benefit (Greene and Hart, 1996). However, any substantial improvement in the description of a collection's sources may lead to better criteria on which to reduce its redundancy, and to set priorities for its maintenance and performance trials (Steiner and Greene 1996).

Apart from decisions on which area and taxa to target for field collection, the issues of efficiency and more effective sampling design can be addressed by map analysis. Traditionally there has been an emphasis on maximizing the number of accessions brought home, glutting the system with an often limited regard to the materials' representation of the targeted adaptation or the diversity of the landscape. Mapping helps by providing an overview which leads to more diversity from fewer samples: this is achieved in part by eco-geographic stratification and in part by examining off-route opportunities to reduce the ever-present bias of windshield survey. Maps also provide a medium for ranking collecting opportunities by uniqueness and ease of access. Thus the use of models and their map derivatives can rechannel expedition energy towards a more strategic and potentially rich set of samples.

In general the use of empirical models helps create a region-oriented deductive check on what has been todate a largely site-oriented inductive search process. By using "adaptive interpolation" and associated methods (Daly et al 1994, Hart 1988), local estimates of habitat character can be calculated more precisely using an array of sources, not limited to the direct assignment of the nearest known data loci. The map domain of GIS extends these model applications by providing a medium for interpolation, so that parameters at measured sites can be applied intelligently to the unmeasured sites of germplasm collection, especially if there is a supporting (or "adaptive") factor such as elevation already in continuous mapped form. It should be mentioned that plant habitats which yield important genetic material are frequently divergent inliers within broader mappable spatial units, so that map data alone are usually too coarse to capture all the habitats' key ecological attributes. However, the dangers of oversimplification are outweighed by the benefits of multi-attribute field verification and a systematic option for developing dynamic sampling plans. Field observations strengthen the foundation which maps provide by yielding detail which the maps miss, and by redefining in experiential terms the true character of the map legends.

Turning Concept into Practice

Dataset Construction

The Caucasus team made a major effort (especially Russian members) to provide useful thematic data from both published and non-published sources. After months of design discussions, searches and data evaluations, the following information was collected: Landsat TM and MSS imagery for identification of key vegetation units, soils maps for edaphic attributes, road maps for access, terrain maps of elevation, slope angle and aspect, and a large volume of climate data.

The prime requirement of the model and map preparations was that field scientists be able to assimilate the results with some confidence regarding the manner in which the data were derived. Care was needed to ensure that the collectors were not alienated by the forced adoption of a new set of tools. The challenge was to show them opportunities to improve their traditional effectiveness, without a distracting amount of extra effort or steep learning curves on their part. Under these circumstances a complex set of algorythms for the data processing would have strained the trust of the collector-user, lengthening the field time required for acceptance.

Accordingly, the methods were as straightforward as possible: raw input data was used without weeding, and mapped to a UTM system of nested cells, so that raster data of differing cellsize would fit together in aggregate when needed. Terrain data were mapped as elevation to 500 meter UTM gridcells (Figure 1.), and as slope and aspect (Figure 2.) to 250 meter UTM gridcells. It is the elevation raster data (or DEM) which formed the base for the climate interpolations.

Five attributes of longterm monthly climate (precipitation (Figure 3.), humidity, maximum and minimum temperature, (Figure 4.) and average windspeed) were mapped from over 100 stations at an average interval of 30 kilometers, and were corrected for elevation via a three-part process. First, normalization of the station data for each month used multiple linear regression coefficients from station easterliness, northerliness, and elevation. Then the normalized station array was kriged to create a trend surface as if the study area were flat and sitting at a median elevation. Thirdly, the elevation effect was restored for every raster in the continuous map by using the DEM in combination with the regression coefficients in reverse.

As an initial reduction of the plethora of 60 monthly climate measures for each 500 meter cell, zones were mapped for moisture (Figure 5.) and for temperature (Figure 6.) by ISODATA clustering on sets of nine and six seasonal combinations respectively. In each result the spatial pattern of clusters showed robust and clear behavior for both amounts and seasonality.

The satellite imagery was rectified to 62.5 (MSS) and 31.25 (TM) meter pixels on a UTM projection, and was presented with grid reference and date labels, but no other obscuring annotation. Thus the prime targets of meadow patches could be evaluated for stable phenology through the multidate dimensions of seasons and years. These data were also the best local reference for familiarizing the team with the potential for collecting in neighboring areas on travel timescales of minutes and hours.

Soil polygons were digitized from four distinct map series which varied in scale, date, geographic fidelity and legend specificity. Each legend entry was given a unique numeric code, and the polygons were plotted with these numbers as labels to guide field markup. The lack of agreement between the different series was a concern from the beginning, as was the inconsistency of map unit size and shape. Roads and major trails were taken from 1950s US 1/250,000 series, and were updated with more recent Russian maps at smaller scales. Satellite imagery was examined to refine alignments where there were questionable features.

The original concept of reducing the entire set of GIS maps to a Terrain Unit stratification turned out to be too ambitious. The all-important soils data were not sufficiently reliable, the climate attributes were best thresholded in a customized manner for groups of plants with similar tolerances, and the terrain variability was on a frequency which would generate tens of thousands of Terrain Unit polygons. There was skepticism that such a combination would result in a useful reduction of the huge amount of cross-variability within the study area, and in the end the fixed strata would only have been explanatory in the manner of a least common denominator. It was decided to keep habitat attributes accessible in their original parametric distributions.

The field "kit" from the map preparations consisted of five two-sheet series of mounted prints of the climate zones, terrain data, and an image mosaic, along with road and soils overlays. These were augmented by a folio of 80 medium-format prints of satellite scenes at a more local scale.

Field Experiences

Prior to the team's convening in the field, a set of map slides were distributed with orientation legends so that each member could preview the data formats and the overall general study area patterns. The attitude towards spatial thematic data at the beginning was generally of high interest but with some concern about its routine use. It was apparent early on that the maps could have been designed with easier legends, and that their linkages to such source data as climate station position and monthly totals should have been more direct.

The maps immediately catalyzed a review of the need for sampling on a regular distance interval, and over several days a consensus formed that the slopes of ecogeographic gradients should determine sample frequency. Whether the maps documented such gradients was a question of concern during a "test period" of about ten days duration. In this period the career-long habits of intensive, inductive collecting strategies were challenged by the deductive indicators of the maps. But the consistently strong corroborations between map and landscape caused a general increase in confidence through the first week, to the degree that the material was adopted as the primary medium for determining potential sites 2-3 days in advance.

It was not, however, a case of technology taking over. The map indicators did not supplant the discussion and decision-making based on direct observation. And while the maps were appreciated for their successful summary of conditions over broad areas, they were obviously inappropriate for predicting habitat subtleties at the level of the sampled plant populations. It was necessary and often surprising to fill out detailed site characterization forms. However, the mutual support of observations and maps led the team members to use each medium to better benefit, and it is in this subtle synergy that the maps made perhaps their most worthwhile contribution. The presence of another authority made one's eyes search for more definite clues, as if there were a heightened need to prove one's observational conclusion. It did turn out that many samples were taken off-route and in peculiar places which probably would have been overlooked had the maps not been used. While it is too early to say if these sites contributed to a broader sample of genetic variability, it was an underlying assumption which provided the impetus for making troublesome sidetrips.

There were other practical uses of the maps. For example, a series of sites were designated as gradient test arrays to form experimental sets for comparing habitat variability and genetic variability. Another strategic use was as an accounting tool to track which habitats had been sampled for a taxon. Unsampled habitats within the species' probable range were then put on a site priority list. After several attempts to find these omissions, they were deemed improbable and dropped off the list with more confidence than if the map accounts had not been available. Lastly, with the knowledge that the maps had already effectively documented certain attributes of the sites (eg. elevation, macroclimate, insolation), the team members could concentrate on those observations which were only possible on site (eg. soil pH, drainage, land use influences).

Overall the acceptance of the maps' utility by the field team was surprisingly complete, given that the scales of geography and of montane habitat ecology are to a large degree non-overlapping. The correspondence of the two perceptual realms would have been less if the modeling approach had not produced the refined local estimates of longterm climatic conditions. The success of scale transferability was stimulating in the general sense of team energies, and in the specific cases of two key scientific discourses: firstly on the strategies for sampling genetic variability, and secondly on exploring the relationships between the broadly-described habitat factors and the very narrow and localized samples of genetic composition.

Tangible Results and Conclusions

Evaluation of Seed Samples

A longterm benefit of the model/map project is to provide continuous attribute surfaces which can be queried at either established or potential seed sampling sites (Figure 7.). From these queries it has been possible to build a set of eighty climate and terrain attributes for each seed sample - including the temperature and moisture regimes for the site at the time of most common germination, midgrowth, maturation, and overwintering. This information is a fresh empowerment of considerable value to the understanding and utilization of the crop genetic material gathered on the expedition.

Perhaps more importantly, the spatial data developed in this project will support experiments into the linkages between genetic and environmental variability. Some taxa exhibit greater phenotypic plasticity and less differentiation in measurable genetic terms. Other taxa form large numbers of distinct genotypes whose distribution has little correspondence to habitat change. Still other taxa show measurable genetic differences which do correlate to the habitats from which they were collected. It is of significance to quantify the degree to which of these possibilities apply to the major taxa collected during the expedition. For the cases where genetics do correspond strongly to collecting site habitat, the research can evaluate the relative strength of linkages among the habitat attributes, and in quantitative terms may offer a calculable model of the environmental factors associated with each sampled adaptation.

Having explored the linkages on a broad variety of forage legume germplasm, the scientists in both countries will be in a better position to set priorities for attaching such map-based descriptors to historic samples within their collections. Finally, the reporting of success and failure in the search for these linkages will affect the objectives and the planning for future efforts to acquire wild germplasm adapted to specific environmental conditions, particularly for crop types related to those tested.

Mapped Results as a Knowledge Base?

How well will the mapping results stand as a general ecological reference? With respect to reliability, the spatial geometry of both maps and imagery compared quite well with the 100+ GPS points established at the collecting sites. Thematic data success was concentrated in the terrain and climate categories, where field patterns generally matched the maps. Roads were out of date, and were both incomplete and overly optimistic, while the four distinct soils compilations proved grossly misleading on too many of the field spot checks.

In terms of non-quantitative assessment, the climate maps are certainly more accurate in those areas close to stations with no intervening topographic barriers, and are probably better for those months where the elevation influence was strong. The weaker map areas were apparent during the field itinerary, but for this application the production of somewhat flawed continuous maps was not a waste of effort, because even for those areas the study did produce characterization which forced the team into local learning once the data weaknesses were exposed. It is fortunate that the continuous nature of the raster maps allows a flexible selection of areas for update, such that pockets where geometries or theme results are problematic can be recalculated with either new or reselected inputs.

Verification of the climate surfaces was not performed. Perhaps the only means to check the behavior of the longterm monthly aggregates is to withhold stations from the primary generation of the surfaces, in order to use them as verification data. This was not done on this project, largely because the needs for precise characterization of local habitats outweighed doubts about the subtleties of the surfacing process. In retrospect this may have been an overly expedient decision, but it was difficult to justify effort towards the assessment of supporting spatial data, when the expedition's primary objectives were in the areas of botany, plant ecology, and population genetics.

As stated in other presentations at this meeting, the use of models and standard map processing should render project results open to general use, with the expectation that the protocols are repeatable and are familiar enough to be accepted broadly by future users of the database. Often a model construction or a GIS compilation contains special adjustments for the initial case study, so that inconsistencies arise when follow-up activities include other applications within the same geography or a reprocessing into adjacent or different areas. Standardization is a boon to reuse of initial results, but carries with it a cost in terms of flexibility and ease of update.

There are two fundamental reasons why ecogeographic mapping initiatives will never be completely done. First the world is a changing place. Its environmental dynamics are described only in part by the historic periods of record. Secondly, the source material for map projects of the type described in this paper are always going to be deficient in some regard, and usually will have only partial improvement during any five year period of update. For these reasons the data themselves should be evaluated on the basis of whether they are good enough for now, as opposed to whether they are good enough forever. Similarly the use of models and GIS protocols should be judged against today's state-of-the-art, with the expectation that in five years we will have the benefit of much more method development and hopefully at least an equal amount of learning from mistakes. It is in meetings such as this current one that the community of practitioners can update their sense of today's best options in terms of tools, tricks and the raw material of parametric data.

Acknowledgements

We would like to thank the field team members who contributed to an extraordinary set of discussions on the expedition: Mel Rumbaugh, Nikolay Khitrov, and Walter Graves. Stephen DeGloria and David Bates of Cornell University also gave valuable inputs. Lastly the data processing was supported by the diligent help given by Stephen Smith, Chris Duncan, Anja Preylowski and Bill Brown.

References

Bolton, J.L., Goplen, B.P. and Baenziger, H. (1972) World distribution and historic developments. *In* Hanson, C.H. (ed.) *Alfalfa Science and Technology*. Agronomy 15:1-34.

Brown, A.H.D., and Marshall, D.R.(1995) A basic sampling strategy: theory and practice. *In* L. Guarino, V.R. Rao, and R. Reid (eds.) *Collecting Plant Genetic Diversity: Technical Guidelines*. Cab International, UK.

Daly, C., Neilson, R.P., and Phillips, D.L. (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J Applied Meteorology* 33:140-155.

Greene, S.L., and Hart, T.C. (1996) Plant genetic resource collections: an opportunity for the evolution of global data sets. *Proceedings of the Third International Conference on Integrating GIS and Environmental Modeling.* NCGIA Santa Barbara, WWW and CD.

Hart, T.C. (1988) Upper Jubba Watershed Performance (JESS Report No. 38), Ministry of National Planning and Jubba Valley Development, Mogadishu, Somalia.

National Research Council (1993) *Managing global genetic resources: agricultural crop issues and policies.* National Academic Press, Washington D.C.

Reid, R. and Strickland, R.W. (1983) Forage plant collection in practice. *In* Mclvor, J.G. and Bray, R.A. (eds) *Genetic Resources for Forage Plants*, CSIRO, Australia.

Steiner, J.J. and Greene, S.L. (1996) Proposed ecological descriptors and their utility for plant germplasm collections. *Crop Science* 36:(in press).

Author Information

Thomas Hart
Spatial Data Associates
20 Bush Lane, Ithaca, NY 14850
Phone: 607-257-0951
Fax: 607-257-3167

e-mail: jhart@lightlink.com

Stephanie L. Greene
Forage Legume Curator, USDA, ARS National Plant Germplasm System
Washington State University-Prosser
24106 N. Bunn Road, Prosser, WA 99350
Phone: 509-786-9265
Fax: 509-786-9370
e-mail: sgreene@ars-grin.gov

Alexandr Afonin
Agricultural Ecologist, N.I. Vavilov Plant Industry Institute
Bolshaya Morskaya, 42, St. Petersburg, Russia, 190000
Phone:812-465-0751
Fax: 07-812-311-8762
E-mail: alex@dma.spb.ru

Ecological modelling in GIS

Peter van Horssen
Department of Environmental Studies
University of Utrecht, Utrecht, The Netherlands
email : P.vanHorssen@frw.ruu.nl

Abstract

An existing non-spatial logistic regression model (Barendregt, 1993) has been modified and implemented in a GIS as a step towards landscape ecological modeling of a wetland area. The model can calculate responses for 102 plant species in marsh ecosystems to spatially variable environmental conditions. Input variables derive from existing maps or from interpolation between point measurements. The output is a map of ecological groups. These maps are validated by records on absence and presence of species of these ecological groups.

Introduction

In the Netherlands interventions in the surface water and ground water system have affected water quality and water budget, to the detriment of the plant species composition of marsh land ecosystems The main activities are ground water extraction for industrial and drinking water supply and lowering of the surface water levels for agriculture in winter and spring. Water shortages in summer and late spring are compensated by intake of nutrient rich surface water from a nearby river system. This water infiltrates and causes ground water quality to change to nutrient rich water types (Wassen 1990, Barendregt 1993, Schot, 1991).

The project aims to select scenarios, by means of a landscape ecological model, for sustainable use of land and water including surviving wetland ecosystems. An integrated landscape ecological has been build to study the combined effect of interferences in this aquatic and phreatic system on the quality of marsh ecosystems. The model integrates the biotic and abiotic component and the spatial and temporal aspects of the system and focuses on water and nutrient transfers by process-oriented modeling of ground water and surface water flows. Environmental site conditions are computed and taken as input variables for the vegetation response model. Because the landscape ecological model must work in an spatial manner, all input and output variables are processed in a GIS but before this research the vegetation response model was not spatially dimensioned. This study describes the implementation of this model (Barendregt, 1994) in a GIS (PCRaster)(Wesseling en van Deursen 1994) and show some results and applications. The study area was chosen in the Vecht streek on its availability of relevant data from earlier research.

Area description

The investigated area (Fig 1) in the central part of the Netherlands contains a polder area with a soil of clay and peat and adjacent a sandy glacial ridge. The polders lie between +1.0 and -3.5 above mean sea level (a.m.s.l.). Surface water levels are completely controlled on fixed winter levels and lower summer levels. The land use is predominately agricultural, mainly dairy cattle breeding, nevertheless substantial parts are natural reserves with lakes, broads, marshes and peat land. The glacial ridge (+1 to + 30 a.m.s.l.) has coarse sandy soils and gravels. The dominant landuse is forest (both decidious and coniferous), heathland and towns (population 50.000 - 75.000). Several groundwater pumping stations are situated here.

Data

The vegetation response model was developed on the basis of 906 record of vegetation and environmental conditions in the investigation area (Barendregt & Wassen ,1989). The model is based on gaussian logistic regression technique and predicts a response value related to the values of 18 different environmental variables per plant species. This particular model calculates the response values of 102 plant species and uses information on soil texture, management, phreatic conditions and ground water quality. Data are used from the period 1985 - 1990. The ground water quality data (Fig 2 and (Fig 3) are processed into maps by means of spatial interpolation resulting in input maps (Fig 4 and Fig 5) for ground water quality with estimated block mean averages of 500 x 500 x 10 meter blocks. Data for validation are taken from independent environmental inventories. The model has been validated using a resolution of 1 x 1 km.

Results

The model yields species maps (Fig 6) which show the response value per plant species to the environmental conditions. For the presence or absence of a species a cut-off level of *0.5 * the maximum calculated response value* is taken. Is a response below this level, a plant species is considered to be not present in the cell, if above the species is considered to be present in the cell. Species are classified according to the ecological groups of (Den Held c.s.1992). The three distinctive properties are : (Appendix 1)

- nutrient status
- water type
- species composition.

A division in nutrient status classes only shows a dominance for plant species related to ground water poor in nutrients. This pattern (Fig 7) is consistent with the expectation that the `higher' grounds show vegetation types of nutrient poor situations whereas areas close to the Vecht river show vegetations of more nutrient rich water types. Mesotrophic vegetations types are found in a topographical intermediate position. A division in water type classes only shows a dominance for plant species related to poikilotrophic water types. This pattern (Fig 8) shows great similarities with the pattern connected with nutrient state classification.

The groups of species compositions are ordered in decreasing chemical loads concentrations. So the A B C E group species are generally related to nutrient rich brackisch/groundwater like water types, while J K L group species are related to nutrient poor rainwaterlike water types. The species of the D F G species group are composed of species of both intermediate nutrient and mixed rainwater - groundwater like water type. Species of group j, f and c are respectively the most found species groups. (Fig 9)

Validation

All valid estimates, that is the found versus calculated and the not found versus not calculated cases, are summed in a valid map (Fig 10) while all the non valid cases are summed in a non valid map (Fig 11). The pattern of the valid map shows a small part of the area in the southwest and south with valid predictions > 80 cases. Most of the area shows valid predictions between 40 and 60 cases,while an area with relatively few valid predictions (< 20) is found in the west central part of the area. The map with the non valid cases shows (logically) the complementary pattern of the map with valid cases.

Discussion & Conclusions

A vegetation response model has been integrated in a GIS and the results in coherent and valid patterns at the landscape scale. The validity of the model was tested with input variables which are estimated from measurement by means of interpolation. Errors caused by interpolation of the input data and in the actual regression model affect the results. In the final model, where the input data are generated by process models, the error in the results is expected to be lower (this remains to be investigated). This constructed spatial vegetation response model can be used as a tool to asses the impact on wetland ecosystem.

Appendix 1

Classification of plant species (Den Held,1992)

- -nutrient status:
    - Eutrophic : nutrient rich
    - Mesotrophic : intermediate nutrient richness
    - Oligotrophic : nutrient poor
- -water type:
    - Atmotrophic : chemical composition rain water like
    - Poikilotrophic : intermediate between atmotrophic and lithotrophic
    - Lithotrophic : chemical composition groundwater like
    - Glyphotrophic I : intermediate between lithotrophic and thalassotrophic
    - Thalassotrophic : chemical composition seawater like
- -species composition.
    - A species : eutrophic nutrient state and a mix of lithotrophic/glypotrophic water type
    - B species : eutrophic nutrient state and a mix of lithotrophic/glypotrophic water type
    - C species : eutrophic nutrient state and a mix of lithotrophic/glypotrophic water type
    - D species : mesotrophic nutrient state and a poikilotrophic water type
    - E species : eutrophic nutrient state and glypotrophic water type
    - F species : eutrophic/mesotrophioc nutrient state and poikilotrophic water type
    - G species : mesotrophic nutrient state and a poikilotrophic water type
    - H species : oligotrophic nutrient state and a poikilotrophic water type
    - K species : oligotrophic nutrient state and a atmothrophic water type
    - L species : oligotrophic nutrient state and a atmothrophic water type

back to text

Literature

Barendregt A. (1993) *Hydro-ecology of the Dutch polder landscape. thesis, Faculty Geographical Studies, University of Utrecht, Utrecht, pp. 200.*

Barendregt, A. en Wassen, M.J. (1989) *Het hydro-ecologisch model ichors (versie 2.0 en 3.0). report Department of Environmental Studies, University of Utrecht, pp. 72.*

Den Held,A.J.,Schmitz, M. & Wirdum van G. (1992) *Types of terestrializing fen vegetation in the Netherlands in: Verhoeven, J.T.A. (eds) Fens and Bogs in the Netherlands Vegetation dynamics, history, nutrientdynamics and conservation, pp 237 - 321. Kluwer Academic Press. The Netherlands.*

Schot, P.P(1991) *Solute transport by groundwater flow to wetland ecosystems. thesis, Faculty of Geographical Studies, University of Utrecht, Utrecht. pp 134.*

Wassen, M. (1990) *Water flow as a major landscape ecological factor in fen development. thesis, Faculty Geographical Studies, University of Utrecht, Utrecht, pp. 197.*

Wesseling, C. and van Deursen, W. (1993) *The PCRaster package. Department of Physical Geography, University of Utrecht.*

Peter van Horssen
Department of Environmental Studies
University of Utrecht, Utrecht
The Netherlands
P.vanHorssen@frw.ruu.nl

Christopher S. McGarry

Geologic Modeling for Landfill Screening: Integrating GIS with Geospatial Modeling

ABSTRACT

The Illinois State Geological Survey is conducting a study to map and model the geology of Carroll County, Illinois to assist county officials as they address landfill development issues. This paper presents the process and preliminary results from map and model analysis. The geologic setting includes thin glacial drift overlying faulted, gently dipping bedrock strata consisting of Silurian dolomite and Ordovician shale and dolomite. The dolomite units serve as regional aquifers; therefore, areas underlain by shallow dolomite were considered poorly suited for future landfill development. Areas underlain by shale were preferred because the shale has a relatively low hydraulic conductivity, which should inhibit potential contaminant migration. Geologically optimal areas for constructing a landfill are characterized by fine-grained glacial deposits (for use as daily cover material) overlying thick, unfaulted shale.

The mapping and modeling process used ARC/INFO geographic information system (GIS) software and earthVision geospatial modeling software. The GIS was used to compile well data, create a digital topographic surface, and generate cross-sections for subsurface lithologic and stratigraphic analysis. EarthVision was used to construct a bedrock topographic surface and geologic unit isopachs using lithologic and elevation data from the project well database. EarthVision was also used to generate a three-dimensional model depicting the geologic framework of the county. Information from the model was compiled using the GIS to produce a map depicting the potential for groundwater contamination from future landfill development.

INTRODUCTION

The Illinois State Geological Survey (ISGS), in cooperation with the Illinois Department of Commerce and Community Affairs and Carroll County, is conducting a study to map and model the shallow geologic framework of Carroll County, Illinois. The project objective is to produce a suite of maps to help county officials address geologic questions concerning capable sites for landfill development. The maps include surface topography, thickness of Quaternary deposits, bedrock geology and geologic capability for landfill siting.

Identifying areas suited for waste disposal facilities has become an increasingly important issue in many areas because existing landfills are reaching capacity (Illinois Environmental Protection Agency 1994). To identify potential new sites for disposal facilities, the geologic materials must be carefully mapped and categorized according to physical properties affecting hydraulic conductivity. In Illinois, the fine-grained Quaternary glacial sediments (i.e. silt and clay) and selected bedrock units (i.e. shale) constitute the geologic materials exhibiting relatively low hydraulic conductivities. Developing landfills in these relatively impermeable materials minimizes the potential for contamination of groundwater resources. Areas underlain by thick deposits of these materials are more suitable for hosting solid waste disposal facilities than areas underlain by materials having a higher hydraulic conductivity. This paper presents a summary of the preliminary analysis performed for Carroll County. Final map products will be produced by Fall, 1996, after careful review.

Geologic Setting of Carroll County

The geologic units commonly penetrated by water wells and other shallow bore holes in Carroll County include Paleozoic bedrock deposited as marine sediments overlain by unlithified Quaternary sediments, predominantly glacial deposits. Larson et al. (1993) provide a stratigraphic and hydrostratigraphic column. The lithologic and hydrologic properties of these units are discussed below.

Ordovician System

The oldest bedrock units exposed in Carroll County are the St. Peter Sandstone and Glenwood Formation of the Ancell Group of Ordovician age. The St. Peter Sandstone is a pure, chiefly medium grained, very well sorted quartz sandstone and is generally 100 to 200 feet thick in this area. The Glenwood Formation consists of approximately 25 feet of interbedded dolomite, fine to medium grained sandstone and shale (Willman et al. 1975). The Ancell Group outcrops along the eastern edge of the county. The high hydraulic conductivity, large thickness and vast extent of the sandstone units of the Ancell Group make them a large potential groundwater resource (Visocky et al. 1985).

Overlying the Ancell Group are the Platteville and Galena Groups, which consist of a continuous sequence of relatively pure dolomite ranging in thickness from 0 (where removed by erosion) to approximately 300 feet in the southwest part of the county (Willman and Kolata 1978). These units are extensively jointed (Foote 1982). These fractures create a secondary porosity and act as conduits for groundwater flow. Much of the groundwater extracted in the northern third of Illinois comes from these jointed dolomite units (Visocky et al. 1985).

The Maquoketa Shale Group is the uppermost Ordovician unit present in Carroll county. It ranges in thickness from 0 to about 200 feet and is characterized by fractured silty, dolomitic greenish-gray shale with argillaceous dolomite lenses (Kolata and Graese 1983). Although the unit is fractured, the joints have collapsed due to the fissile nature of the shale. Where present, the low hydraulic conductivity of the Maquoketa Shale Group inhibits groundwater flow and contaminant migration. It has been referred to as the Maquoketa Confining Unit in regional aquifer studies (Visocky et al. 1985, Larson et al. 1993).

Silurian System

The formations of the Silurian System form the uppermost bedrock in the northern and southern part of Carroll County. These rocks consist of fractured and jointed relatively pure dolomite; some formations contain argillaceous or cherty beds, which were not differentiated for this study. They are differentiated by Willman (1973). The total thickness of the Silurian rocks ranges in thickness from 0 to approximately 150 feet in Carroll County (Willman et al. 1975, Larson et al. 1993). Fractures within the Silurian formations permit groundwater flow and subsequent dolomite dissolution, producing caves and sinkholes. Visocky et al. (1985) identified the Silurian formations as a primary aquifer of the Upper Bedrock Aquigroup in northwestern Illinois. Larson et al. (1993) stated that this aquifer "can be highly susceptible to contamination where it occurs close to the ground surface or is overlain by coarse grained material."

Quaternary System

The glacial deposits of Carroll County are complex in geometry and variable in composition. A sandy diamicton, the Ogle Till, covers the bedrock in the southeastern part of the county. Extensive lacustrine deposits, classified as the Equality Formation, are found in lowland valleys in the west-central part of the county. The Peoria Loess, a wind blown silt, covers the entire county (Willman and Frye 1970). Other features include numerous kames and sand dunes found in the southeastern part of the county.

Geologic Structure

In addition to the extensive jointing of all bedrock units throughout the county, the northern part of the county has undergone post-Silurian faulting. The half-mile wide Plum River Fault Zone is a near vertical fault downthrown to the north with strata displacement of 100 to 400 feet (Kolata and Bushbach 1976). Other features include the Upton's Cave Syncline north of the Plum River Fault Zone and two structural domes south of the fault zone on the eastern edge of the county, the Forreston Dome and the Brookville Dome (Treworgy 1981).

Geologic Landfill Screening Criteria

Berg et al. (1984) rated areas with unfractured shale overlain by glacial till as exhibiting a low potential for aquifer contamination and cited the Maquoketa Shale as a primary representative, stating "these bedrock materials provide the highest order of natural protection." Consequently, mapped areas outside the Plum River Fault Zone where thick Maquoketa Shale is overlain by more than 50 feet of glacial till were designated as having the greatest potential for landfill development in Carroll County.

MODELING AND MAPPING TECHNIQUE

The ISGS well database for Carroll County and a one mile buffer around the county contains records for 1220 wells, including private and municipal water wells, structure and highway borings, and exploratory wells. These well records were imported into geographic information system software, ARC/INFO by Environmental Systems Research Institute, Inc., for cross-section construction, stratigraphic analysis and spatial representation. In addition, ARC/INFO was used to generate an accurate digital topographic surface from U.S. Geological Survey (USGS) 7.5-minute topographic quadrangle maps. Geospatial modeling software, earthVision by Dynamic Graphics, Inc., was used to construct contour maps of the surface topography and the thickness of Quaternary deposits and a three-dimensional model of the lithologic and stratigraphic framework of the geology of the county from data imported from the project well database. The contours for the map of the thickness of Quaternary deposits, the bedrock geology contacts and relationships among subsurface geologic units were extracted from the model and compiled using the GIS to produce a landfill capability map.

Surface Topography

An accurate digital representation of surface topography was needed to create a realistic three- dimensional model of the geologic units in Carroll County. Surface topography data were compiled using Digital Raster Graphic (DRG) files from the USGS Mid-Continent Mapping Center and ArcScan, a component of ARC/INFO. DRG files are pseudo-color images scanned from 7.5-minute topographic quadrangles at a resolution of 8.2 feet per image pixel. Each DRG file was georeferenced and transferred to an ARC/INFO grid format. ArcScan, raster-to-vector conversion software, was used to trace and attribute topographic lines from the grid and create an ARC/INFO coverage of the surface topography for each quadrangle in the county at a 1:24,000 scale. The line coverages were transformed to point coverages containing data representing the topographic lines. The topographic point data were imported and gridded in earthVision using a grid spacing of approximately 150 ft. x 110 ft. to create a grid for recontouring at a 50 foot interval. A very fine grid spacing was chosen to accurately represent the vast amount of input data. The contour lines generated in earthVision were exported back into the GIS for coverage integration and map output generation (fig. 1).

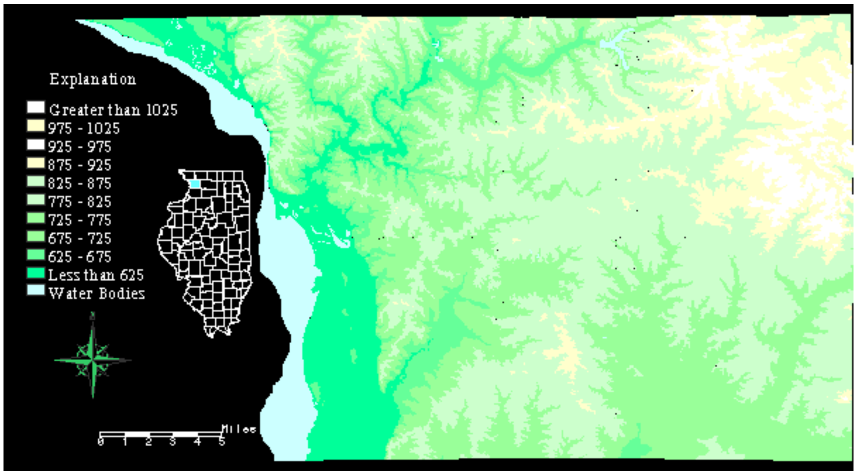


Figure 1

Surface Topography in Carroll County, Illinois. Elevations are in feet above mean sea level. Contour interval is 50 feet.

Well Database

Well data were divided into two categories: header information (data pertaining to the well location, well type, driller's name, date drilled, etc.) and geologic unit information (data pertaining to each unit's lithology, thickness, depth to top of the unit, etc.). This database design permits independent treatment of the data in map view and cross-section.

Extensive topographic relief in the county made it necessary to manually determine well locations by visual inspection of USGS 7.5-minute topographic quadrangle maps. Locations were based upon the Public Land Survey System information (section quarters, section, township and range) and property ownership plat books. Approximately 21% of well locations were plat book verified. Elevations of geologic units within a well were calculated from the ground surface elevation; therefore, accurate well locations were crucial.

Upon review of the spatial distribution of well data, three clusters of data were recognized. Two clusters reveal areas of higher residential development and one cluster represents an area previously studied as a potential nuclear power plant site. Wells within these areas comprise approximately 39% of the well database, yet represent only 6% of the total area within the county. Therefore, a large portion of the county contains few or no data points to support the maps produced for this project.

The accuracy of the lithologic descriptions of the geologic units in the well database was tested by constructing geologic cross-sections. Cross-sectional views of all wells within a one mile wide traverse in east-west and north-south directions were generated with an in-house program. Each distinct unit described in the driller's well logs was categorized by lithology (e.g. dolomite, shale, sandstone or fine or coarse grained Quaternary deposits). Geologic units were assigned group and/or formation names according to stratigraphic correlation of similar lithologies. Well records containing lithologic descriptions not conforming to the lithologies of surrounding wells were discarded. The anomalous well records may be a result of poor driller's logs or incorrect well locations.

Thickness of Quaternary Deposits

The thickness and lithology of Quaternary deposits are important factors in the landfill screening process. Waste disposal facilities are constructed in these near-surface geologic units and these materials are used as landfill cover. For Carroll County, the thickness of the Quaternary deposits was derived using software tools to compare the surface topography to the bedrock topography. Using the project well database, the uppermost bedrock unit elevations were identified and gridded with a grid spacing of approximately 925 ft. x 775 ft. This grid spacing was found to most accurately represent the bedrock surface with the available data. The bedrock surface grid was subtracted from the surface topography elevation grid to generate a thickness of Quaternary deposits grid. This grid was contoured at a 50 foot interval and the contours exported into the GIS for analysis. Lack of bedrock data within topographically rugged areas where the bedrock is near or at the surface, such as the northwest portion of the county, inhibited a realistic representation of the bedrock topography. The contour map produced by the earthVision software depicted areas of thicker Quaternary deposits where thinner deposits were known to exist. These map features were edited and/or deleted in the GIS to more accurately represent the Quaternary deposit thickness. Relatively little area contains more than 50 feet of Quaternary deposits (fig. 2).

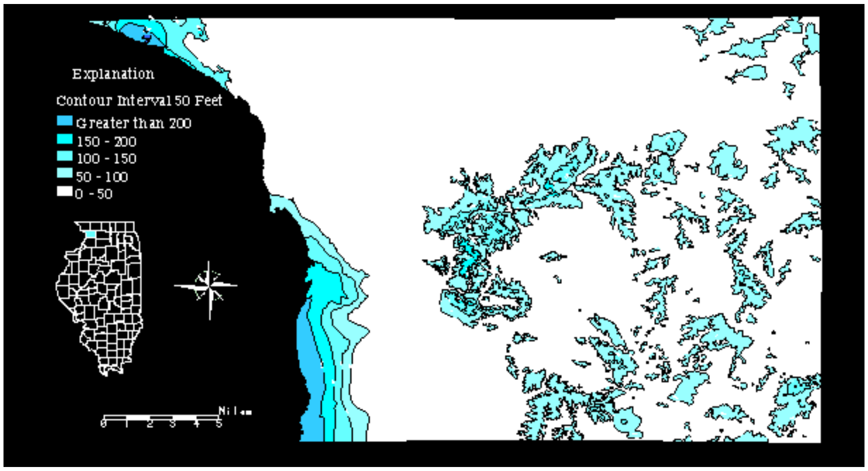
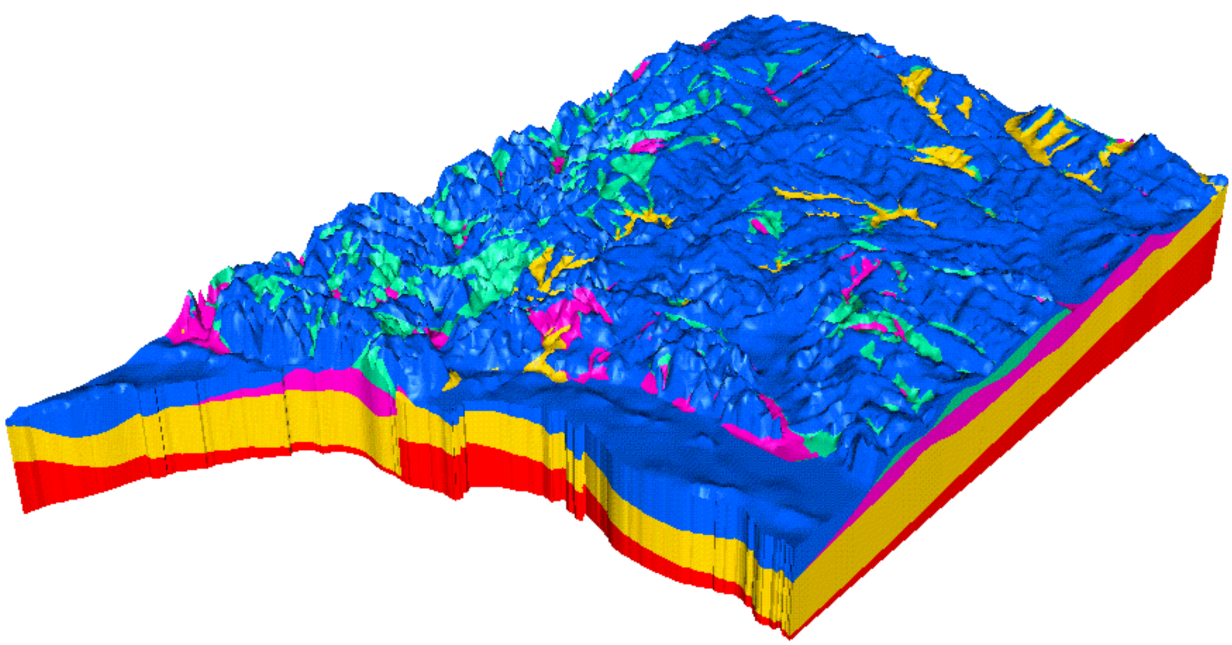


Figure 2

Preliminary Map of the Thickness of Quaternary Deposits in Carroll County, Illinois. Contour interval is 50 feet.

Bedrock Geology

Elevation data for each bedrock unit were extracted from the well database, separated according to their location with respect to the Plum River Fault Zone, gridded with a grid spacing of approximately 925 ft. x 775 ft. and integrated into the Geologic Structure Builder component of earthVision. The Geologic Structure Builder creates a three-dimensional representation of faulted areas from two dimensional grid surfaces and three dimensional grid isosurfaces by developing a fault tree. Each branch of the fault tree is separately calculated and juxtaposed to form the three dimensional structure (Dynamic Graphics, Inc. 1994). The fault tree used in this study consisted of two branches, north and south, on either side of the Plum River Fault Zone. Each branch utilizes the respective grids for the region and creates a layer-cake geometry depending upon the type of grid interface chosen. Interface options used include a depositional surface (e.g. the two dimensional grid is deposited on the next lower surface without truncating any underlying suface) and an unconformity (e.g. the two dimensional grid is deposited on the lower surface; if the grid nodes fall below an underlying surface, the lower grids are truncated). The resultant grid spacing chosen for the three-dimensional model was approximately 2400 ft. x 2100 ft. x 40 ft. due to software limitations. The three-dimensional model reveals the displacement of bedrock units in the Plum River Fault Zone, the gentle southwestward dip of the bedrock units and the wide Mississippi River flood plain on the western edge of the county (fig. 3).



Key: Quaternary = blue; undifferentiated Silurian = green; Maquoketa Shale = pink; Galena and Platteville Groups = yellow; and

Ancell Group = red.

Figure 3

Preliminary Model of the Near-Surface (Shallow) Geology of Carroll County, Illinois

Bedrock contacts were extracted from the model and imported into the GIS for display and analysis. The abrupt contact of the undifferentiated Silurian formations with the Galena and Platteville Groups marks the Plum River Fault Zone. Relatively little area is underlain by Maquoketa Shale, one of the units ranked as being more suited for hosting a landfill.

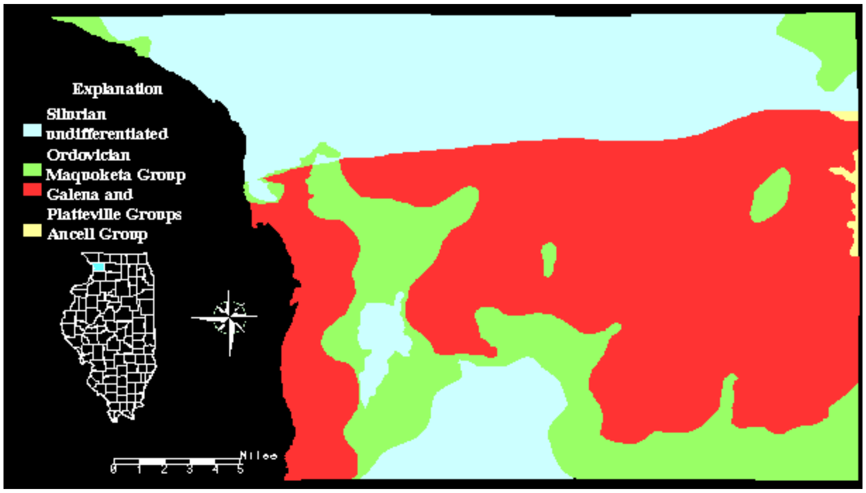


Figure 4

Preliminary Map of the Bedrock Geology of Carroll County, Illinois

Geologic Capability for Landfill Siting

The probable geological capability for hosting a landfill was mapped as a guide to help county officials avoid areas which have little chance of meeting their landfill siting criteria. Final selection of any site requires a site-specific investigation because of the scarcity of data in parts of the county and the map scale.

The classification scheme used by Berg et al. (1984) was simplified into three catagories: high, intermediate and low geologic capability for hosting a landfill. The bedrock geology map was overlain by the thickness of Quaternary deposits map to locate areas where Maquoketa Shale is overlain by more than 50 feet of fine grained Quaternary deposits. Berg et al. (1984) determined that these areas provide excellent groundwater resource protection from landfill leachate. These areas were classified as having high geologic capability for hosting a solid waste disposal facility, assuming waste disposal trenches 20 feet deep. Areas within the Mississippi River Valley were classified as having low geologic capability because the Maquoketa Shale is beneath over one hundred feet of river sediments (primarily sand and gravel) which exhibit a high hydraulic conductivity and would provide little or no protection from the potential migration of contaminants. Areas where the Maquoketa Shale is overlain by less than 50 feet of Quaternary deposits were ranked as having an intermediate geologic capability. Areas not underlain by Maquoketa Shale were ranked as having low geologic capability for hosting a landfill because the dolomite units that overlie and underlie the Maquoketa Shale are regional aquifers capable of rapidly transmitting contaminants (fig. 5).

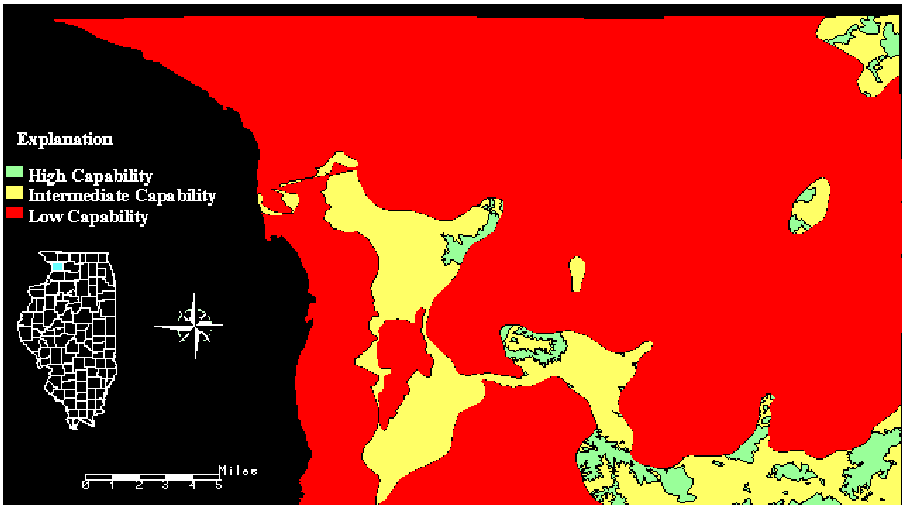


Figure 5

Preliminary Map of the Geologic Capability for Landfill Development in Carroll County, Illinois

Preliminary Map Limitations

The preliminary maps presented in this paper have not yet been checked for accuracy with other regional or state-wide studies. No edge-matching or comparison has been conducted with published maps. Other parameters to be included in the geologic capability of hosting a landfill are the topographic surface slope and the distribution of sand and gravel deposits within the Quaternary sediments. Much additional work remains before final map products can be published.

SUMMARY

The integration of geographic information system software with geospatial modeling software has enabled the Illinois State Geological Survey to create a three-dimensional model of the shallow geology of Carroll County. Geologic unit contacts and contours were extracted from the model and imported into the GIS for display and analysis. Areas characterized by more than 50 feet of fine grained Quaternary deposits underlain by Maquoketa Shale were interpreted as potentially highly capable of hosting a landfill. The shale has a relatively low hydraulic conductivity, thus inhibiting potential contaminant migration, and the Quaternary deposits are suitable for use as daily landfill cover material. Preliminary map products presented here include suface topography, thickness of Quaternary deposits, bedrock geology and geologic capability for landfill siting. These maps were created to assist county officials as they address geologic issues concerning landfill development. In addition, the geologic information compiled and interpreted for this study provides parameters which should be considered in future groundwater protection studies.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the contributions to this project made by Curtis C. Abert, Ryan P. Fairfield, Maureen D. Kelly, Dennis R. Kolata, Robert J. Krumm, E. Donald McKay, Melisa M. McLean and Darrell E. Rennels.

This project has been supported, in part, by the Illinois Department of Commerce and Community Affairs and Carroll County, Illinois.

REFERENCES

Berg, R.C., Kempton, J.P., and Cartwright, K. (1984) *Potential for Contamination of Shallow Aquifers in Illinois*. Illinois State Geological Survey Circular 532.

Dynamic Graphics, Inc. (1994) *EarthVision User's Guide*. Alameda, California: Dynamic Graphics, Inc.

Foote, G.R. (1982) *Fracture Analysis in northeastern Illinois and northern Indiana*. Master's thesis, University of Illinois.

Illinois Environmental Protection Agency (1994) *Available Disposal Capacity for Solid Waste In Illinois, Seventh Annual Report*. Illinois Environmental Protection Agency, Bureau of Land IEPA/BOL/'94-149.

Kolata, D.R., and Buschbach, T.C. (1976) *Plum River Fault Zone of northwestern Illinois*. Illinois State Geological Survey Circular 491.

Kolata, D.R., and Graese, A.M. (1983) *Lithostratigraphy and Depositional Environments of the Maquoketa Group (Ordovician) in Northern Illinois*. Illinois State Geological Survey Circular 528.

Larson, T.H., Graese, A.M. and Orozco, P.G. (1993) *Hydrogeology of the Silurian Dolomite Aquifer in Parts of Northwestern Illinois*. Illinois State Geological Survey Environmental Geology 145.

Treworgy, J.D. (1981) *Structural Features in Illinois- A Compendium.* Illinois State Geological Survey Circular 519, Plate 1.

Visocky, A.P., Sherrill, M.G., and Cartwright, K. (1985) *Geology, Hydrology, and Water Quality of the Cambrian and Ordovician Systems in Northern Illinois*. Illinois State Geological Survey and Illinois State Water Survey Cooperative Groundwater Report 10.

Willman, H.B. and Frye, J.C. (1970) *Pleistocene Stratigraphy in Illinois*. Illinois State Geological Survey Bulletin 94.

Willman, H.B. (1973) *Rock Stratigraphy of the Silurian System in Northeastern and Northwestern Illinois*. Illinois State Geological Survey Circular 479.

Willman, H.B., Atherton, E., Buschbach, T.C., Collinson, C., Frye, J.C., Hopkins, M.E., Lineback, J.A., and Simon, J.A. (1975) *Handbook of Illinois Stratigraphy*. Illinois State Geological Survey Bulletin 95.

Willman, H.B., and Kolata, D.R. (1978) *The Platteville and Galena Groups in Northern Illinois*. Illinois State Geological Survey Circular 502.

Christopher S. McGarry
Assistant Staff Geologist
Illinois State Geological Survey
615 East Peabody Drive
Champaign, Illinois 61820
Telephone: (217) 244-7413
FAX: (217) 333-2830
E-mail: mcgarry@bert.isgs.uiuc.edu

Groundwater Monitoring in the Alluvial Aquifer of the River Sieg, Germany - An Application of MODFLOW/MODPATH combined with GIS Analysis

Christian Michl
Friedrich-Schiller-University Jena
Department of Geography
Loebdergraben 32
D-07743 Jena, Germany
Tel: 03641-632776, Fax: 03641-632727, e-mail: c5chmi@rz.uni-jena.de

Abstract

Geographic Information Systems (GIS) have been increasingly applied for geohydrologic research studies as they provide a variety of spatial analysis tools for groundwater modeling. In this study the GIS IDRISI was linked to the modulary structured groundwater model MODFLOW to simulate the hydrological dynamic of the alluvial aquifer of the River Sieg, Germany. Using a 10-year groundwater data base imported into the GIS IDRISI the hydraulic conductivity and the transmissivity of the aquifer were analyzed by raster overlay techniques and imported into PROCESSING MODFLOW and MODFLOW. After establishing the hydraulics of the aquifer, MODFLOW was calibrated for steady-state and transient conditions. Additionally, the particle tracking model MODPATH was used to detect conceptual errors in the assumed boundary conditions, and in the simulated infiltration from the rivers into the aquifer. MODFLOW simulated the groundwater dynamics in the aquifer with a good fit to the observed data for steady-state and transient conditions. By using the GIS raster data base and PROCESSING MODFLOW an effective and improved pre- and post-processing was obtained. Future research will focus on the linkages (i) of the river dynamics. (ii) of the unsaturated zone with the aqiufer, and (iii) of the MODFLOW model with the GIS GRASS.

Ling Bian, Hao Sun, Clayton Blodgett, Stephen Egbert, WeiPing Li, LiMei Ran, Antonis Koussis

AN INTEGRATED INTERFACE SYSTEM TO COUPLE THE SWAT MODEL AND ARC/INFO

ABSTRACT

This project presents results of an interface development that couples GIS (Arc/Info) and an advanced hydrologic model SWAT (Soil and Water Assessment Tool). The SWAT model is designed to predict water and sediment yields for large rural river basins; it has been applied to many major river basins in the United States with promising results. However, extracting spatial parameters and entering a great number of parameters as required by SWAT are tedious processes. The interface system developed in this project is to (1) streamline the GIS processes for preparing spatial parameters required by SWAT, (2) automate the link between Arc/Info and SWAT, and (3) provide a user-friendly data entry and editing environment for the SWAT model. Because both Arc/Info and SWAT are mature and complex systems, the interface, therefore, includes add-on external graphical user interfaces and an object-oriented internal database to couple the two systems. The external graphical user interfaces facilitate spatial parameter extraction and data entry. The object-oriented database is well suited for linking the two systems. Hydrologic components were treated as different classes of objects, and the associations among them were coded to identify internal inconsistency, adding intelligence to data entry and editing.

INTRODUCTION

Presently, the major impediment to progress in hydrologic modeling is not the inability to simulate physical processes mathematically, or to solve the resulting equations, but the inability to explicitly consider the spatial variation of model parameters (Maidment 1993). This task has always been one of the most time consuming, and therefore costly, components of hydrologic modeling. Spatially distributed data and the ability to manipulate such data are essential for detailed hydrologic modeling. In the 90s, hydrologic modeling concentrates on developing model "front ends" (pre-processing of input data to prepare parameters for physically-based models) and "back ends" (post-processing of results visually, statistically, and numerically).

GIS offers precisely the capabilities to account for the spatial variability of hydrologic processes. Although hydrologic parameters can be generated within GIS, the procedures are not sufficiently streamlined to be handled by non-GIS specialists. In many cases, the linkage between GIS and hydrologic models still remains a manual operation. These problems have become a major hurdle in the current effort to enhance the spatial capabilities of hydrologic modeling using GIS. Currently, many GIS data appropriate for water resources management are underutilized due to the absence of a bridge between the raw data and end users.

OBJECTIVES

The need for integrating GIS and hydrology is well understood; the link itself, however, is weak at the operational level and is incompatible with the demands of hydrologic modeling. The primary objective of this study is to develop an interface system to integrate Arc/Info, a widely used comprehensive GIS software package, and SWAT (Soil and Water Assessment Tool, Arnold et al. 1993), a hydrologic model of advanced capabilities. The specific tasks are to provide the following capabilities: (1) to streamline GIS processes tailored toward hydrologic modeling needs; (2) to automate data communication between GIS and the hydrologic model; and (3) to provide a user-friendly data entry and editing environment for the hydrologic model.

METHODOLOGY

The Hydrologic Model

The hydrologic model, SWAT, is a semi-empirical and semi-physical model. It has been used as a practical model to predict the effect of agricultural management decisions on water and sediment yields for large ungauged rural watersheds. Moreover, SWAT is an advanced lumped model or a semi-distributed model because it allows a watershed to be divided into hundreds of areal units, either by polygon sub-watersheds or regular grids. This semi-distributed characteristic is well suited for integration with GIS.

SWAT consists of major water budget components such as weather, surface runoff, return flow, percolation, evapotranspiration, transmission losses, pond and reservoir storage, crop growth, irrigation water transfer, groundwater flow, and channel routing. The model runs on a daily time step for short or long term predictions and operates in a semi-distributed manner to account for spatial differences in soils, land use, crops, topography, channel morphology, and weather conditions. Using GIS data, SWAT has been applied to many major river systems in the United States with promising results (Koussis et al. 1994, Srinivasan et al. 1993). The SWAT model has become increasingly popular in recent years.

As an advanced lumped model, SWAT accounts for spatial variability at the expense of data input. SWAT runs as a batch process; a large number of parameter files must be prepared prior to model execution. At the basin level, SWAT can take more than ten separate input files concerning agricultural management, water bodies, basin configuration, and weather information. At the subbasin level, SWAT can use up to nine input files containing detailed information for subbasin characteristics, surface and ground water bodies, channel routing, soils, weather, and agricultural practices. Using SWAT for a ten-subbasin watershed, a user may have to prepare nearly one hundred input files, with each containing ten to thirty parameters. SWAT provides a rather primitive user interface to facilitate data entry and editing and operates only in the DOS environment. All input files are fixed formatted and must be prepared separately. A more integrated, advanced user interface would significantly enhance the capability and usability of SWAT.

Interface Design

A range of approaches has been proposed and implemented for integrating GIS and environmental models (Abel et al. 1994, Maidment 1993, Chou and Ding 1992, Nyerges 1992). Abel et al. (1994) summarized three major integration architectures. A simple two-component architecture allows for one-way data transfer between two independent systems (e.g., a GIS and an environmental model). It promises low cost of implementation but low usability as well. An "embedded" two-component architecture extends capabilities of a master component by using functions of an embedded agent component. This is a more integrated approach, with the usability and costs depending on the capabilities of the master component. A many-component architecture consists two or more master components that share common agent components such as a database management system and/or an end-user interface. This option provides a single external schema for the integrated system yet retains the independence of each master component. The costs of this architecture tend to be high but it is desirable when the component systems are complex.

The two systems being integrated, Arc/Info and SWAT, are mature and complex, each with its own data model and conceptual and external schema for handling database and user interface tasks. In a simple two-component architecture, streamlined GIS procedures are possible and GIS data can be transferred to SWAT; however, it does not alleviate the burden of processing the large number of individual input files. A second scenario, the "embedded" architecture, is virtually impractical in this case. Implementing functions of one of the systems into the framework of the other would involve substantial modifications for one or both systems. This is costly given the complexity of both Arc/Info and SWAT. On this premise, the many-component architecture is adopted for this study. The two systems, Arc/Info and SWAT, are dealt with as two independent master components in the integration system. The conceptual design of the integration system includes an add-on external user interface and a shared internal database to couple the two systems (Figure 1). Supporting hydrologic modeling is the primary function of the interface system; thus the design intends to accommodate the requirements of the SWAT modeling with GIS playing a supporting role.

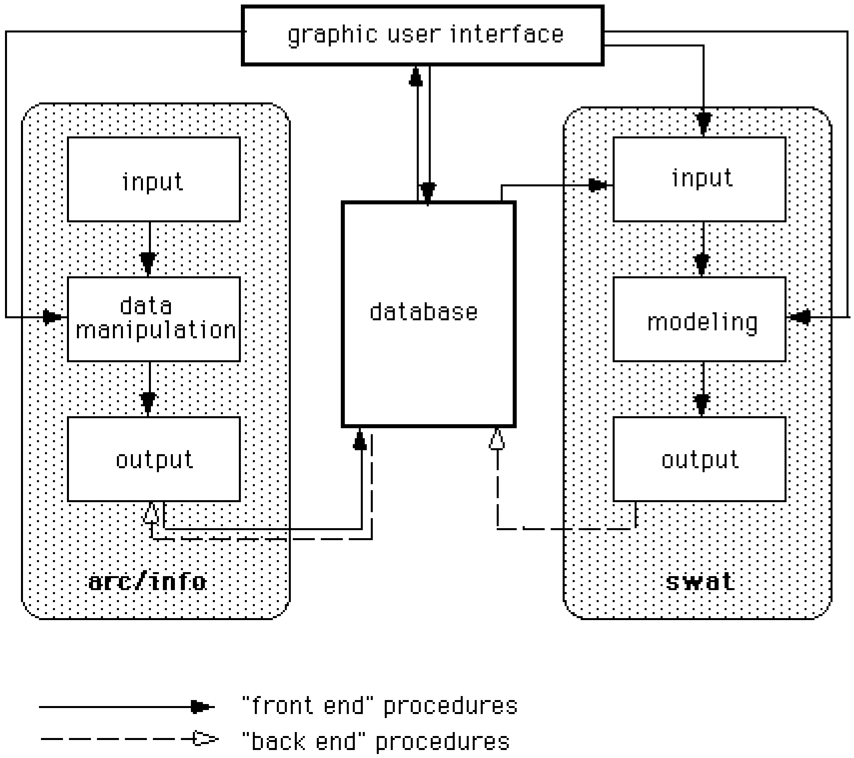


figure 1. architecture of the interface system coupling arc/info and swat.

The integration begins with the external user interface, where the end-user initiates a new database or activates an existing one. AMLs (Arc Macro Language) are activated via the interface to prepare input parameters for SWAT in the Arc/Info environment. The data transition from GIS to the SWAT model is automated through the internal data base shared by both the GIS and the hydrologic model. User-friendly data entry and editing is part of the functionality of the external graphic user interface (GUI), where users can interactively enter and modify model input files and parameters, including non-spatial parameters. The internal database stores the input data and transfers the data into a SWAT compatible format. As the last step, the execution of SWAT is activated through the external user interface. The following sections describe major functions of the AMLs, the GUI, and the internal database.

Spatial Parameters Extraction

The primary function of the AMLs is to streamline GIS processes in order to extract spatial input parameters for SWAT. The AMLs are written in such a way that users with minimal GIS experience can easily operate the interface.

The AMLs use source GIS data that are commonly available. The base data sets include streams, hydrologic code unit maps, digital elevation models (DEM), hyposography maps, temperature, precipitation, soils, and administrative boundaries. All the data involved can be obtained from public domain sources such as US Geologic Survey (USGS), Natural Resource Conservation Service (NRCS), and National Climate Data Center (NCDC).

A majority of the spatially explicit parameters required by SWAT are included in six input files: Basin, Subbasin, Precipitation, Temperature, Soils, and Routing. The extraction process is thus organized in six corresponding routines to ensure compatibility with the SWAT file structure that is explicit to end users; this is also compatible with object definitions in the internal database, thus making data passing efficient. During the operation, the user has the option to process the entire basin or selected subbasins. In addition, the user can terminate, proceed, or at times skip a routine as the program allows. An error checking mechanism warns the user whenever proper procedures are violated, while a "help" function provides detailed instructions for operation.

The Basin, Subbasin, and Routing routines extract spatial information such as basin or subbasin area, basin terrain, and channel morphology. Most parameter extractions are automated processes, while those spatial parameters that require users' familiarity with the basin are secured through an interactive mode. For example, users identify interactively on screen the longest channel in a subbasin by selecting a series of stream segments. An error checking mechanism validates the continuity of the selected channel to eliminate possible errors induced during manual selection. The interactive procedure engages users to the spatial data and avoids the complex, lengthy process of automated search.

Terrain parameters, i.e. average slope length and average slope steepness, are calculated using a method described by Williams and Berndt (1976):

$l = 0.5 * DA/LCH$ (1); $S = 0.25 * Z (LC25 + LC50 + LC75) / DA$ (2)

where l is the average slope length; DA is the area of the subbasin; LCH is the total length of channels in the subbasin; S is the average slope steepness; Z is the elevation range in the subbasin; and LC25, LC50, and LC75 are contour lines generated at 25, 50, and 75 percent of the total elevation range in each subbasin, respectively.

Using GIS functions, the temperature and precipitation data at point locations are interpolated into Thiesson polygons. The values of the parameters are averaged over all polygons in a subbasin and weighted by their areal contributions to the subbasin. Because of the spatial scale SWAT is designed for large rural river basins, the State Soils Geographic (STATSGO) database is used as the base data. Soil parameters are weighted by both soil horizon depths and areal percentages of soil series in a soil association. Users have the option to keep or discard any soil series.

Graphic User Interface (GUI)

The GUI is a Windows-based tool that communicates with the internal database for data entry, editing, query, data validation, and launching the SWAT program. It is designed to overcome many of the most significant hurdles faced by SWAT users, particularly the lack of a user-friendly interface and the tedious input files processing. The development environment for the GUI is Visual Basic Professional operating under Microsoft Windows. Visual Basic is a Rapid Application Development (RAD) tool for interactively creating Windows programs. The GUI, in combination with the internal database, interfaces with Arc/Info and PC SWAT.

The GUI follows the standard Windows single-document interface (SDI) format common to many Windows programs and employs standard Windows controls (menus, frames, command buttons, and radio buttons, etc.). A pull-down menu is used as the basic form for the interface. It best accommodates the existing structure of SWAT input files and parameters. In addition, it provides the user with a familiar graphical user environment that requires minimal training. The operational functions of the interface are organized in a hierarchical structure, where menu items are the top level controls which evoke client windows or dialogue boxes for interactive data processing.

At the top menu level, the GUI allows the user to initiate a new SWAT project or open an existing one for modification. In either case, the user is presented with standard Windows dialogues to guide the process. The interface also provides the user an option to import spatial information extracted from the Arc/Info environment or enter the parameters directly. Data entry, editing, and query are the main functions of the GUI. Once these are complete, the user may save the data to the internal database and execute the SWAT model.

Data entry is divided into two main groups: basin and subbasin. Since many of the forms are similar between the two groups of menus, they are color-coded to provide a visual cue for the user. Both the master basin and the master subbasin forms consist of an array of command buttons, one for each input file type. By selecting a specific input file type, the user can access a data entry form that contains a number of text entry boxes for individual parameters. Alongside each parameter box is the name of the parameter and a brief description of the parameter. Each box is keyed to the data compilation sheets that accompany the SWAT documentation. The user may switch at any time across subbasins and basin, or change between basin (or subbasin), file type, or parameters.

For many SWAT parameters, the user needs to estimate the values based on the hydrologic conditions of the basin. SWAT documentation provides reference tables to assist the user in choosing appropriate values. These are implemented in the GUI as dynamic lookup tables that are accessible by selecting the text entry box for the parameter of interest. The values displayed in the lookup tables can be easily adopted and saved to the internal database. Parameters that have multiple values in a time series (i.e. one value for each month of the year) or multiple components (e.g., a list of pesticides) are entered into a two dimensional spreadsheet-like grid. This allows the user to examine all the values for the parameter at once. Similar to the individual grids, a Subbasin Summary function displays all subbasin data in multiple tables, with each table containing all parameters of a specific input file type across all subbasins. This helps the user to evaluate and compare the parameter values over the entire basin. The remaining file types are organized as separate tables behind the one currently being displayed; the user can easily switch file types by selecting file labels.

Internal Database

The internal database is developed to automate data communication between Arc/Info and SWAT. Through the external graphical user interface, the internal database imports spatial information extracted in Arc/Info and supports all database functions such as data entry, editing, and query. It interfaces with SWAT by transferring input data into required SWAT input files. The database is developed in an object-oriented approach using C++.

Development of the database requires an object-oriented requirement analysis. Typically, the requirement analysis consists of three basic components: an object model, a dynamic model, and a functional model. In the particular case of this database, the object model is the primary concern, while the dynamic model and the functional model are more relevant to hydrologic model development.

The basis of the object model is objects and their relationships, which can be represented by classification, inheritance, association, and aggregation (Khoshafian 1993, Kainz and Shahriari 1993, Milne, et al. 1993, Egenhofer and Frank 1992, Oosterom and Bos 1989). Figure 2 presents an overview of the conceptual design of the database. Each rectangle represents an object while all linear symbols depict how the objects are organized in the database. The arrows indicate the classification hierarchy, in which the superclasses are at the points while the contributing subclasses are at the tails. The straight lines depict association or aggregation relationships among objects. Lines with nodes on both ends indicate associations; objects at both ends are members of an association set. Aggregation is represented by lines with a node and a diamond at either end; objects at the node sides are components of the composite objects adjacent to the diamonds.

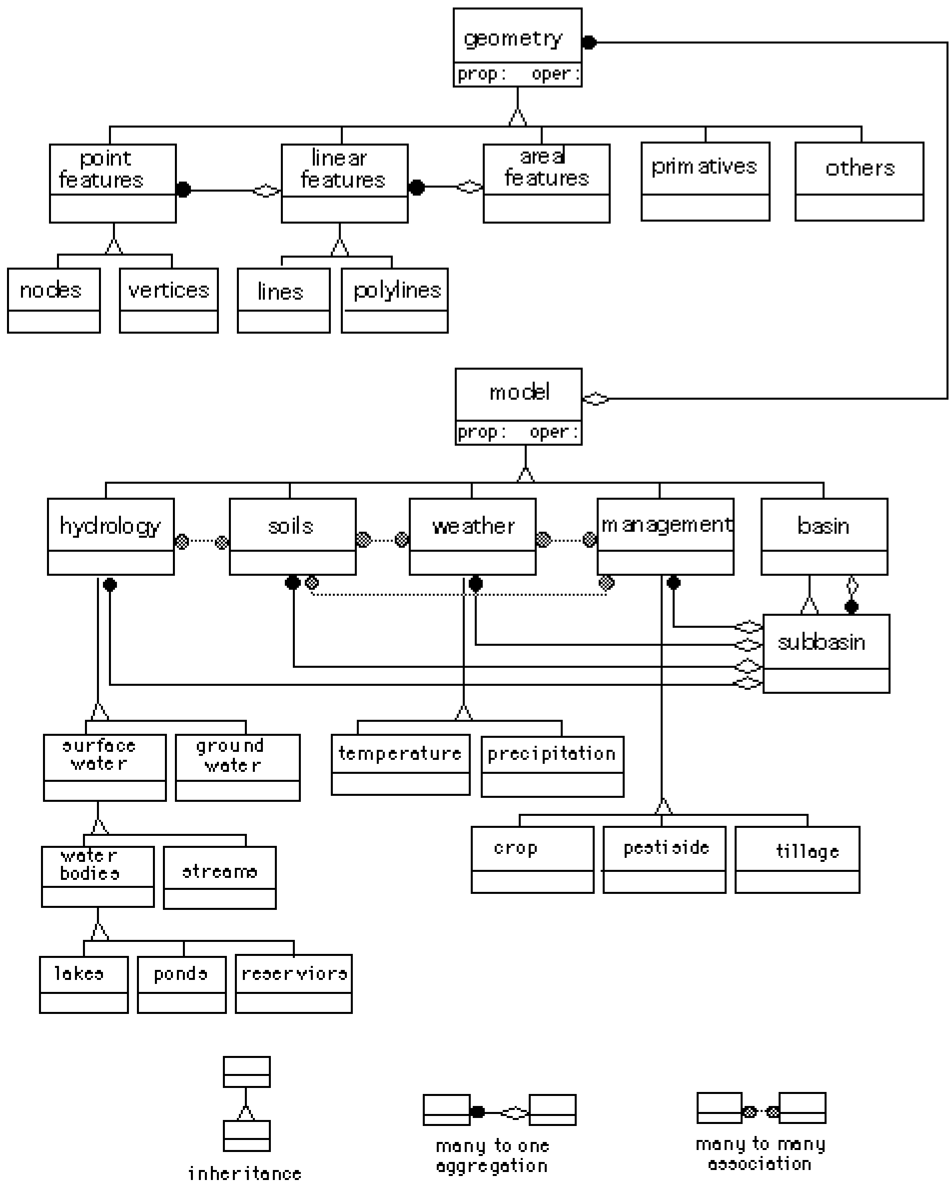


figure 2. the typeview of the data model for the internal database.

There are two main object types, GEOMETRY and MODEL. The spatial objects and parameter objects are dealt with as two distinct objects types, instead of treating parameters as properties of the spatial objects. This is well suited to the design principle of the interface because hydrologic modeling is the primary concern. The design of the object model should accommodate the requirement of the hydrologic model.

The GEOMETRY object type is used primarily to handle spatial information derived from Arc/Info. The actual calculation of areal weighting for spatial parameters of soil and weather data is carried out by the internal database, while Arc/Info only provides intermediate data such as the area of individual Thiessen polygons and areas of subbasins. This object type includes various geometric objects such as point, line, and area features; each of these is in turn a superclass of more specialized spatial objects, which are subclasses of their superclasses. All geometric objects share common operations such as computing percentage and length or area. The point and linear objects are related to linear and areal objects, respectively, in a multiple-component aggregation relationship (Figure 2). Specifically, point objects are components of its composite object: linear object, and the linear objects are components of the areal object.

The MODEL objects are organized as a hierarchy of hydrologic parameters. The objects are so defined that they closely correspond to the real world entities. In the mean time, the definition is closely related but not restricted to the file structure of SWAT, in which many input files correspond to real world entities such as soils, water bodies, weather, and agricultural practices. As part of the model, the GEOMETRY objects are aggregated into the MODEL object.

In the database, individual parameters are treated as properties of the primitive objects, which are the most specialized objects and no longer decomposable. For example, all soils parameters are the properties of primitive object SOIL. The object SOIL inherits all properties and operations of MODEL (e.g., maximum and minimum value checking) in the database and have parameters and operations unique to soils. A particular soil series is an instance of SOIL. As part of a subbasin, most the subclasses of the MODEL object are aggregated to SUBBASIN object, which in turn was aggregated to the BASIN object.

The associations among the subclasses of the MODEL object are implemented according to hydrologic principles that are actually implemented in the SWAT model. Typically, hydrology, weather, soils, and agricultural management are related to each other in simulating hydrologic processes. Implementing such associations validates internally input parameter values. For example, water stress factor is computed by considering evapotranspiration, which is in turn calculated using temperature (Figure 3). Water stress factor can also be entered independently, whose value can be evaluated according to the association and temperature input. This is a more advanced validation mechanism than those available in many other database environments, for example, the relational database. In addition, this validation mechanism can be easily implemented in an object-oriented environment.

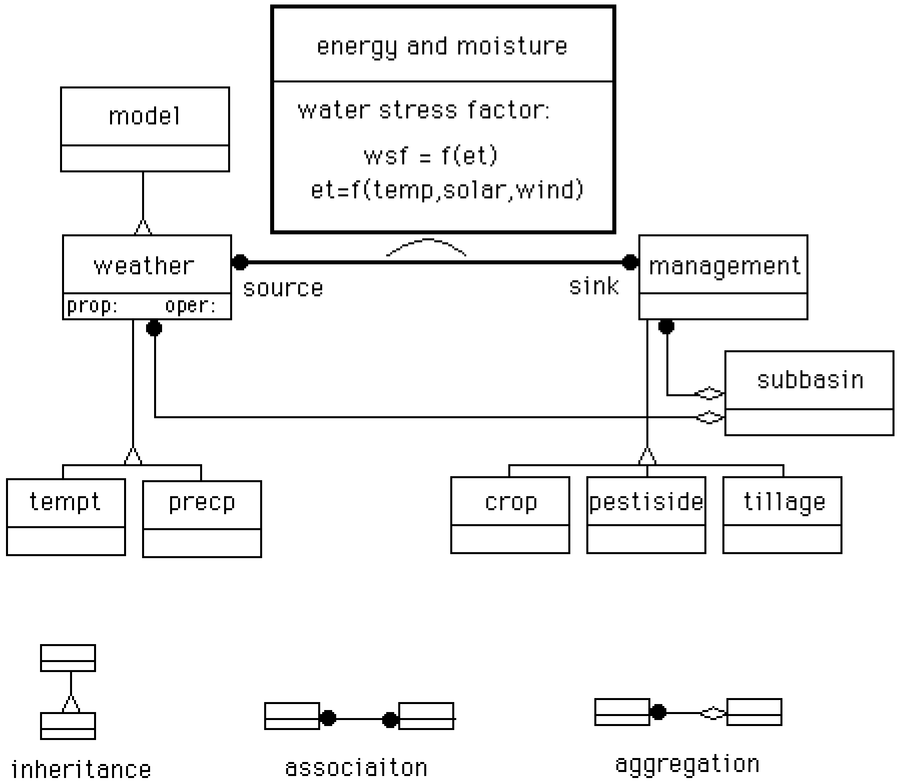


figure 3. association between weather and management.

The database is implemented in Dynamic Linked Library (DLL) format. An Application Programming Interface (API) is developed to communicate between the GUI and the internal database. Through the API, all operations received at the user interface can be implemented in the database. Once all the data entry, editing, and query are completed, the user activates a save function from the GUI; the internal database updates the input data and transfers them into SWAT ready files.

CONCLUSION

The interface system described above will allow hydrologists to fully exploit the capabilities of SWAT and GIS systems. With this interface, large amount of GIS data will become readily usable for the hydrologic modeling community. It is also hoped that this research will contribute to the better understanding of integrating hydrologic models and GIS, and ultimately contribute to effective water resources management.

ACKNOWLEDGMENTS

This work was funded by the Kansas Water Office and the University of Kansas General Research Fund. We are grateful to Dr. Pedro J. Restrepo for his expert opinion in the design of the interface.

REFERENCES

Abel, D.J., Kilby, P.J. and Davis, J.R. (1994) The systems integration problem. International Journal of Geographical Information Systems, 8(1):1-12.

Arnold, J.G., Allen, P.M., and Bernhardt, G. (1993) A comprehensive surface groundwater flow model. Journal of Hydrology, 142: 47-69.

Chou, H.-C., and Ding, Y. (1992) Methodology of integrating spatial analysis/modeling and GIS. Proceedings, 5th International Symposium on Spatial Data handling, Charleston, South Carolina, 514-523.

Egenhofer, M. J., and Frank, A.U. (1992) Object-oriented modeling for GIS. URISA Journal 4(2):3-19.

Kainz, W., and Shahriari, N. (1993) Object-oriented tools for designing topographic databases. Proceedings, GIS/LIS '93, 341-350.

Khoshafian, S. (1993) Object Oriented Databases. John Wiley, New York.

Koussis, A. D., Sophocleous, M., Bian, L., and Zou, S. (1994) Lower republican River Basin: Stream-Aquifer Study, Technical Report, University of Kansas, USA.

Maidment, D. R. (1993) GIS and hydrologic modeling. in Environmental Modeling with GIS, Goodchild, M. F., B. O. Parks, and L. T. Steyaert (ed.) Oxford University Press, New York.

Milne, P., Milton, S., and Smith, J.L. (1993) Geographic object-oriented databases - a case study. International Journal of Geographical Information Systems, 7(1):39-55.

Nyerges, T. (1993) Understanding the scope of GIS: its relationship to environmental modeling. in Environmental Modeling with GIS, Goodchild, M. F., B. O. Parks, and L. T. Steyaert (ed.) Oxford University Press, New York.

Oosterom, P. V., and Bos, J.V.D. (1989) An object-oriented approach to the design of geographic information systems. Computer and Graphics, 13(4):409-418.

Srinivasan, R., Arnold, J., Rosenthal, W., and Muttiah, R.S. (1993) Hydrologic modeling of Texas Gulf Basin using GIS. Proceedings, Second International Conference on Integrating GIS and Environmental Modeling, Breckenridge, Colorado.

Williams, J. R., and Berndt, H.D. (1976) Determining the universal soil loss equation's length-slope factor for watershed. In Soil Erosion: Prediction and Control, the proceedings of a National Conference on soil Erosion, 217-225.

AUTHOR INFORMATION

Ling Bian
Department of Geography
State University of New York
Buffalo, NY 14261-0023
lbian@geog.buffalo.edu

Hao Sun, Leica Inc.
23868 Hawthorne Boulevard
Torrrance, CA 90505

Clayton Blodgett
Department of Geography
213 Lindley Hall, University of Kansas
Lawrence, Kansas 66045-2121
blodgett@falcon.cc.ukans.edu

Stephen Egbert
Kansas Applied Remote Sensing Program
Nichols Hall
Lawrence, Kansas 66045-2969

WeiPing Li
SAI Software Consultant Inc.
4917 Waters Edge Drive, Raleigh, NC 27606

LiMei Ran
ManTech Environmental Technology Inc.
2 Triangle Drive, Research Triangle Park, NC 27709
lran@taiyang.alm-rtp.com

Antonis Koussis
Institute of Geodynamics
National Observatory of Athens
Athens, Greece

Valentina Krysanova, Dirk-Ingmar Müller-Wohlfeil, Alfred Becker

Mesoscale Integrated Modelling of Hydrology and Water Quality with GIS Interface

Water quality modelling for mesoscale watersheds represents a field where a compromise solution between very complex models with many parameters and inadequate simplified models should be found. The paper includes a description of a new watershed model SWIM which integrates hydrology and water quality. It is based on two previously developed models - SWAT and MATSALU - and is intended for application in mesoscale watersheds with an area of up to 20,000 km2. SWIM includes some modules common to both models, trying to combine their advantages and to avoid overparametrization. A SWAT/GRASS interface is one of the advantages of SWAT, and it is adapted for the use in SWIM. The method of spatial disaggregation is more advanced in MATSALU, and thus it is implemented for SWIM. The model runs under the UNIX environment. Simultaneously to the development of the model, it is being tested in subbasins of the Elbe drainage basin.

---

1. Introduction

Watersheds are important as integrators of many forces, including climate. Their natural boundaries and hierarchical structure represent an appropriate structure for environmental impact analysis and modelling. At present, knowledge about the dynamics of dominant processes in large watersheds is still rather limited due to the extremely complicated character of these processes and the interrelation of many factors of different nature (Beven, 1995). Reliable assessment of nonpoint source (NPS) pollution is one of the problems involved, especially for larger watersheds.

Watershed simulation models represent physical and biochemical processes in a dynamic way. Conceptually, such models describe mathematically water fluxes and associated pollutant fluxes from the land surface and soil profile. Source areas can be categorized in accordance with a distinct land use/land cover and soil type. Dissolved and solid-phase concentrations of chemical compounds can be obtained from lumped modelling of biogeochemical cycling at a source area. These concentrations vary with land cover, soil type, management practices and season of year. Transport (or retention) factors reflect a complex chain of physical and biochemical processes which can affect nutrient movement from a subbasin to the river outlet and must be taken into account.

The complexity of the specific watershed simulation model depends on the temporal and spatial resolution, and on the extent to which important biochemical processes are considered in the model. While there is a certain progress in water runoff simulation for larger basins, erosion and pollutant transport from larger watersheds represents a field where a compromise solution between very complex models with too many parameters and inadequate simplified models should be found.

Previous efforts in watershed modelling were concentrated mainly on developing either continuous-time spatially lumped models or single event spatially distributed models. Most of the models tended to focus on the patch scale or small homogeneous watersheds, where data availability is certainly better (models CREAMS (Knisel, 1980), EPIC (Williams et al., 1984), OPUS (Smith, 1992)). Recent development of deterministic models provides also spatially distributed tools, among them: AGNPS (Young et al., 1989), ANSWERS (Beasley et al., 1980), SWRRB (Arnold et al., 1990), MATSALU (Krysanova et al., 1989a & b), SWAT (Arnold et al., 1993). In a sense, all the models made use of previous approaches. In particular, CREAMS and its components were used as a basis for many further tools. For example, SWRRB is a distributed version of CREAMS, which can be applied to a basin with a maximum of 10 subbasins, and SWAT is an extended and improved version of SWRRB, running simultaneously in several hundred subbasins. However, the general tendency is that the data requirements increase exponentially with the increase in watershed size.

The availability of GIS tools and more powerful computing facilities made it possible to overcome many difficulties and limitations and to develop distributed continuous time models, based on available regional information. While the application of AGNPS and ANSWERS is limited to watersheds of about 200 km2, SWRRB was developed with limited distributed parameter capability to be used in agricultural basins as large as 600-800 km2, MATSALU was applied in a 3,500 km2 rural basin, and SWAT is intended to be applied in watersheds up to 25,000 km2. The SWAT represents a component of the HUMUS project, where it is applied for 350 6-digit hydrologic unit areas in the 18 major river basins in the U.S. (Srinivasan et al., 1993b) All these models are to a certain extent integrated with GIS tools.

In this paper a new watershed model SWIM, which integrates hydrology and water quality, is presented. This model is based on SWAT and MATSALU, and is adapted for the use in European conditions. Simultaneously with development, the model is applied for subbasins of the Elbe drainage basin.

The German part of the Elbe river drainage basin (about 96,000 km2) has been chosen for our regional study due to several reasons. Firstly, it is one of Europe`s largest river basins, situated in Central Europe (80% of the area belong to the former GDR), and sharing similarity in climate and data availability with other European rivers. Secondly, the basin with predominately sandy soils is exposed to comparatively low amount of precipitation, and is characterized by high water demand (both climatic and anthropogenic), all these factors predetermine its high hydrological vulnerability. Thirdly, the Elbe is one of the most heavily contaminated water courses in Europe, due to ineffective sewage water treatment and lack of nonpoint source pollution control (agricultural areas cover about 56% of the total area). The new model is intended to be applicable to other river basins in Europe. To assure this, the input data and parameters are preferably taken from the public domain sources and treated in as universal manner as possible.

2. Model development

2.1 General description

The new model **SWIM** (**S**oil and **W**ater **I**ntegrated **M**odel) is a simulation continuous-time spatially distributed watershed model, based on two previously developed models:

SWAT - (Arnold, Allen, Bernhardt, Srinivasan, Muttiah, Walker, Dyke, 1993, USDA & Texas A&M University), and

MATSALU (Krysanova, Meiner, Roosaare, Vasilyev, 1989, Estonian Ac. Sci.).

SWAT is a continuous-time distributed simulation watershed model. It was developed to predict the effects of alternative management decisions on water, sediment, and chemical yields with reasonable accuracy for ungauged rural basins. The model was developed by modifying the SWRRB (Arnold et al, 1990) and ROTO (Arnold, 1990) models for application to large, complex rural basins. Major changes include: (a) expanding the model to allow simultaneous computations on several hundred subbasins, (b) adding components to simulate groundwater flow, routing transmission losses, and sediment and chemical movement through ponds, reservoirs, and streams. The model operates on a daily time step. A discretization scheme allows subdivision into cells and/or subbasins. SWAT is integrated with the GIS GRASS and a relational data base to extract necessary input parameters.

MATSALU is a system of four simulation models for a mesoscale agricultural watershed and the ecosystem of a sea bay. It was developed for and applied in the Matsalu Bay watershed in Estonia in order to evaluate different management scenarios for eutrophication control of the Bay. Spatial disaggregation is based on the overlay of three maps (elementary subwatersheds with an average area of 10 km2, land use, and soil) to obtain so-called Elementary Areas of Pollution (EAP). The model includes four coupled submodels. At first, water balance is calculated for each EAP with a daily time step to define soil moisture and runoff components (modification of the SCS CN method). After that nutrient balance in soil and nutrient losses from the EAP are estimated. Then water and nutrients are routed in stream flow, and, finally, nutrient cycling in the sea bay ecosystem is calculated. While originally the model was not integrated with a GIS, recently some efforts were made in order to provide integration of chemical submodel with ARC/INFO (Meiner, 1995). Nevertheless, the model is not sufficiently transferable to be used directly in other watersheds.

Why was it necessary to develop a new version of a watershed model? The reason is that no one of the existing models can be used directly for any European watershed larger than 1,000 km2. While AGNPS has a limited scale of application, MATSALU is not sufficiently transferable, and SWAT is currently adapted only for US watersheds (using the specific data sets, particularly soil and weather data bases) and is tested with the monthly time step as a long-term predictor. On the other hand, practically all the mentioned models are in development, none of them is perfect.

For example, it is clear that the current scheme of spatial disaggregation in SWAT (a basin is subdivided into subbasins, and then the dominant soil and land use are used to characterize this subbasin as the lumped parameters) could be and should be improved. So, as an initial step, the spatial disaggregation scheme from MATSALU was introduced into SWAT. The next step was to adjust the model for the use in European conditions, where data availability is different. Currently the model SWIM includes some common modules of the both predecessors: the SCS CN method for surface runoff, the MUSLE approach for erosion, a simplified EPIC approach for crop growth, trying to combine their advantages: the three-level spatial disaggregation scheme and nutrient modules from MATSALU; the GRASS interface, hydrological modules and routing procedure from SWAT, and to avoid over-parametrization.

SWIM integrates weather, hydrology, erosion, crop growth, and nutrients (nitrogen and phosphorus) at the watershed scale. The following hydrological components are included:

component	method
precipitation	real input data or weather generator
snow melt	f(water content of snow, daily max temp.)
evapotranspiration	Pristley-Taylor method for PET, Richie's model for AET
surface runoff	modified SCS CN method
subsurface runoff	storage routing technique for soil
percolation to ground water	storage routing technique for soil
ground water flow	f(ground water recharge)

The following chemical components are included:

component	method
fertilisation	input data (agriculture statistics)
mineralization	f(org. matter, temp., soil water)
plant consumption	f(crop growth, crop type)
denitrification	f(nitrate-N, soil water)
leaching into ground water	f(g-w recharge, soluble nutrient content)
sorption/desorption for P	f(soil type, P content)
transport with surface flow	f(soluble nutrient content, surface runoff)
transport with subsurface flow	f(soluble nutrient content, subsurface runoff)

Transport (or retention) factors are taken into account through averaging (weighted average) of water and nutrient fluxes for heterogeneous subbasins and routing of water, sediments and nutrients in the river flow through transmission losses. The model is intended to be applied in mesoscale watersheds with an area up to 20,000 km2.

If we want to have an efficient model of water quality for the regional scale, an appropriate balance in the model parametrization has to be found. On the one hand, the model has to be detailed enough to account for a large diversity of processes involved. On the other hand, if the model is fully deterministic, its use for large watersheds is limited. Current model development tries to overcome these difficulties. The neural network to extract channel geometry in SWAT, and wide use of weather generators for water quality modelling are only two examples. We suggest implementing one else stochastic procedure in regional-scale models, namely the stochastic allocation of crops for agriculture areas.

Really, it is extremely difficult to obtain current data on crop distribution in some tens to hundreds of subbasins, at least for areas with wide varietys of crops and crop rotations. In our opinion, some stochastic crop generators should be used for that in parallel with existing weather generators, especially for larger basins. The idea is that a limited number of crop rotation schemes can be defined for a certain region, based on expert knowledge. After that the crop generator is applied to distribute the crops in a watershed in accordance with land use, soil, and "crop probability" in the region. This procedure is intended to be included in SWIM.

2.2 Spatial disaggregation and integration with GIS

Several water quality models, including AGNPS, ANSWERS, SWRRB, and SWAT, have been integrated with a GIS (most often ARC/INFO or GRASS) to facilitate the use of spatially distributed data. One major difference in extracting inputs for distributed watershed models is the method of disaggregation to drive the input parameters and submodels. Models like AGNPS or ANSWERS divide the study area into square grids, extract inputs for each grid cell, and apply the model to every cell.

In models like SWRRB or SWAT, a basin is divided into subbasins based on elevation only or on elevation and homogeneity of soil and crop. After that, tools are provided to aggregate inputs at both subbasin level (to extract either the mode or weighted average characteristics for soil and crop in a subbasin) and basin level (to evaluate water and chemical routing and transmission losses) for the model. A default procedure in SWAT is used to aggregate the soil series categories and land use for each subbasin using the mode (dominant) aggregation method.

In the MATSALU model the whole basin (about 3,500 km2) was divided into subbasins of major tributaries, which, in turn, were subdivided into so-called Elementary Watersheds (EW) with an average area of 10 km2. A map of elementary watersheds existed prior to the study. After that, an overlay of land use and soil maps onto the EW map created the Elementary Areas of Pollution - homogeneous plots of land with uniform land cover and soil type in the EW. Such a three-level disaggregation has proven to be quite satisfactory.

According to our experience, spatial disaggregation into elementary units based on natural features (topography, land use, soil types, ground water table) is preferable in comparison with disaggregation based on square grids, as this essentially reduces the number of elementary units and computing time while preserving the accuracy of calculations.

A three-level disaggregation is implemented in the SWIM model for meso-scale basins. It is more complicated than that in SWAT, but it seems to be more reliable. The idea is that a mesoscale basin with an area of 50 to 20,000 km2 is first subdivided into subbasins of reasonable average area (see explanation below). It can be easily done using the *r.watershed* program of GRASS (or any other GIS with similar capabilities), which is applied to a DEM (Digital Elevation Model) of the whole area. Then hydrotops or elementary units are delineated within every subbasin, based on land use and soil types. Normally, a hydrotop is a set of disconnected units in the subbasin. The three-level disaggregation implies 1) basin, 2) subbasins, and 3) hydrotops.

As an example, Fig. 1 demonstrates the scheme of spatial disaggregation for the Elbe drainage basin. We do not intend to apply the model for such a large basin as a whole. That is why, the whole watershed (Fig. 1, e) should be first subdivided into subwatersheds of a reasonable average area. In the case of Elbe the whole drainage basin (96,000 km2) was subdivided into 57 subwatersheds with an average area of 1,700 km2 (Fig.1, d) using *r.watershed* function in GRASS. Then the three-level spatial disaggregation scheme is applied to these subwatersheds (which are later called "basins") (Fig.1, c) - into subbasins (Fig.1, b) and hydrotops inside the subbasins (Fig.1, b and a). After this disaggregation of the basin, an aggregation procedure is used for modelling, starting from the hydrotop level up to the basin level, as indicated by arrows at the Fig. 1. After the verification of results for the Elbe subwatersheds (Fig.1, d), the routing scheme can be applied to combine the outputs for the whole Elbe watershed.

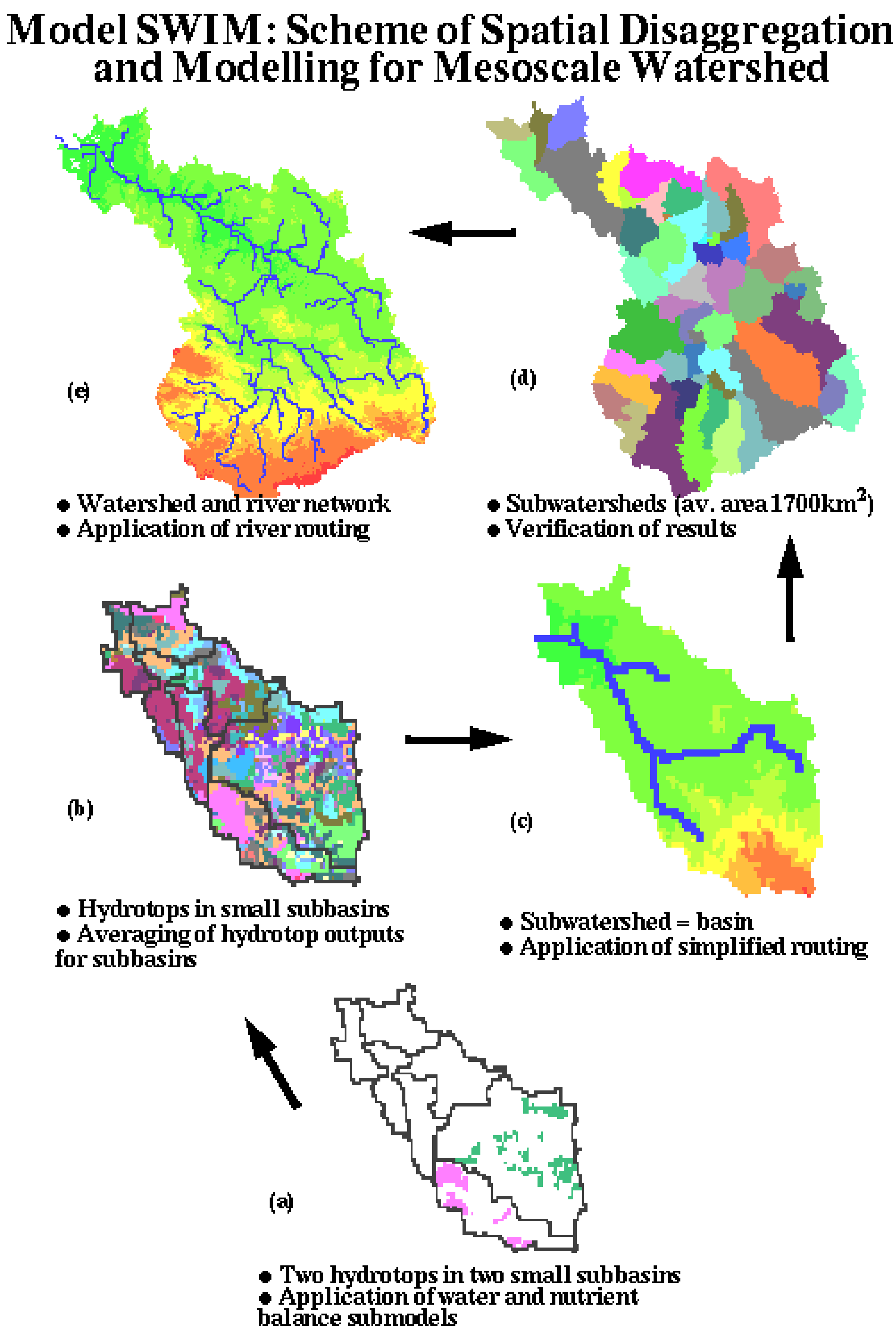


Fig. 1

The SWAT/GRASS interface includes some very useful operations (for example, to extract the routing structure and the weather stations), and it is partially used for SWIM. At the same time, some additional operations were needed to implement a new spatial disaggregation scheme. For example, the SWIM/GRASS interface creates a "structure file" to drive the model. Every line in the file describes the characteristics of one hydrotop - its land use, soil, and subbasin number.

2.3 Reasonable spatial resolution

A very important question is how to choose reasonable spatial resolution? This problem is of fundamental significance for hydrological and hydrochemical process modelling. According to Kuchment (1922), the level of detail adopted in representing the spatial variability of vertical moisture fluxes (which is assumed to be about 100m) depends on the ranges of horizontal mixing in the surface turbulent boundary layer (which should be of the order of 10 times more than this depth). The conclusion follows that it is sufficient to take into account only heterogeneities on scales of 1 km or larger. This means that grid cell size (to be combined into hydrotops) should be not larger than 1 km.

According to Beven and Kirkby (1979), the effect of the channel network probably becomes important for basins larger than about 10 km2, where the time constant of the network (i.e. travel time through it) becomes as long as for the infiltration phase. And it is also known that 10 km (or 100 km2) appears to be the minimum scale, beyond which inhomogeneities in land surface properties can trigger specific meso-scale atmospheric circulation systems, which have a definite impact on land surface - atmosphere interactions (Kuchment, 1992). So, an average subwatershed area, where the effect of the river network can be neglected, should be not more than 10 - 100 km2.

The restriction on average subwatershed area and time step influence the computing time, and, taken together with data availability, define the upper limit of watershed area for the model application. Such reasonable spatial disaggregation should allow the applicability of the model to be extended to larger basins.

3. Model application

3.1 Spatial and relational data

The model operates on a daily time step. The SWAT/GRASS interface (Srinivasan, Arnold, 1993, Srinivasan et al., 1993a) is adopted for SWIM to extract spatially distributed parameters of elevation, land use, soil types, and groundwater table. It is modified where necessary as described below. The interface creates a number of input files for the basin and subbasins, including the hydrotop structure file and the subbasin routing structure file. Steps 1, 2, 5, 6 are used unchanged, steps 3, 4 are new, some other steps from SWAT/GRASS (such as irrigation and nutrient attributes) are excluded.

1. Subbasin attributes. This is the first step to be fulfilled. Using a given subbasin map, the program calculates area, resolution, and coordinate boundaries for the basin and each subbasin. The fraction of each subbasin area to the basin area is calculated.

2. Topographic attributes. The program estimates the stream length, stream slope and

geometrical dimensions, accumulation area, and aspect. The weighted average method is used to estimate the overland slope and slope length. Finally, the channel factors K and C of the Universal Soil Loss Equation (USLE) are estimated using a standard table.

3. Hydrotop structure. The program defines the basin structure by overlaying the subbasin map with land use and soil layers. The structure file is created to run the model. Every line in the file describes the characteristics of one hydrotop - its land use, soil, and subbasin number.

4. Weather attributes. The program selects the closest weather station to the subbasin. Then either actual weather information, or weather generator (in the case the long-term monthly statistics are available for precipitation and temperature for the station) can be used.

5. Ground water attributes. The ground water parameters are estimated for each subbasin using the alpha layer, which defines the time lag needed to the groundwater flow as it leaves the shallow aquifer to return to the stream (Arnold et al., 1993).

6. Routing structure. This very important step in the SWAT/GRASS interface creates the routing structure for subbasins, based on the elevation map. Also, it defines the channel width and depth using a neural network that is embedded in the interface, based on the drainage area and average elevation of a subbasin.

After that the relational meteorological data, crop data base, soil data base, data on point sources of pollution and river routing parameters are read from the files. Additionally, river discharge and concentrations of nutrients are needed for model verification. Currently the model is tested using actual weather data.

The digital soil map of Germany (Bodenübersichtskarte der Bundesrepublik Deutschland 1 : 1 000 000) provides information for 72 soil types. Each soil type is characterized through a "leading profile". For each horizon of every soil profile, 8 attributes are specified: depth, texture class, clay content, humus content, carbon content, nitrogen content, field capacity, available field capacity. Based on the map legends, a special soil data base was created, which is used in SWIM to extract or estimate necessary soil parameters. For example, saturated conductivity is estimated based on the method described in OPUS (Smith, 1992), and soil erodibility factor is estimated from soil texture classes.

3.2 Modelling procedure

After the input parameters are read from files, the three-step modelling procedure is applied:

1. water and nutrient balance are calculated for every hydrotop (Fig. 1 a),

2. the outputs from hydrotops are averaged (weighted average) to estimate the subbasin output, not accounting for lag time in the case of surface runoff, and taking the average for subbasin lag time for subsurface flow (Fig. 1 b),

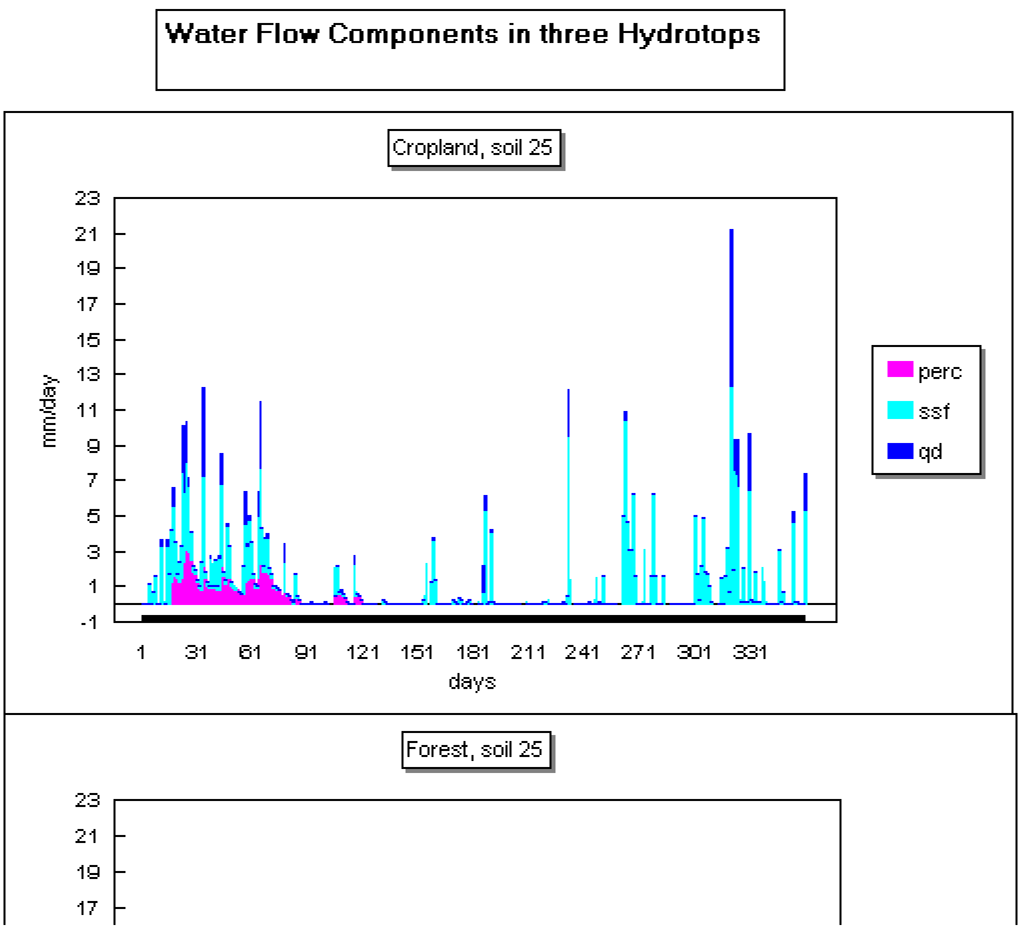
3. routing procedure is applied to the subbasin outputs, taking into account transmission losses (Fig. 1 c).

3.3 Model test

In parallel to the model development, a first application is performed for the Buckener Au catchment (64 km2 - a subbasin of the river Stör and the Dahme basin (up to Markisch Buchholz, about 535 km2) - subbasin of the river Spree (both Stör and Spree are tributaries of the Elbe). These rivers are not regulated and it enables us to test hydrological components with daily time resolution.

Here the results for hydrological cycle modelling are demonstrated for the Buckener Au watershed, which was subdivided into 9 subbasins and 72 hydrotops. Simulation was performed for 4 subsequent years 1989 - 1992.

Time series of different water flow components: surface runof (qd), subsurface flow (ssf) and percolation (perc) in 1990 are shown in Fig. 2 for three hydrotops - cropland on soil 25, forest on soil 25, and pasture on soil 6. The soil 6 is Niedermoorboden in German classification, which corresponds to Eutric Histosols in the FAO classification, the soil 25 is Podsol-Parabraunerde / Podsol-Fahlerde in German classification, which corresponds to Spodic Luvisols/Spodic Podzoluvisols in the FAO classification. The hydrotop "pasture on soil 6" can be considered as a wetland, which is dominated by organic soils. Here, percolation is most intensive (Fig. 2), whereas the surface runoff is highest in the cropland and lowest in the forest.



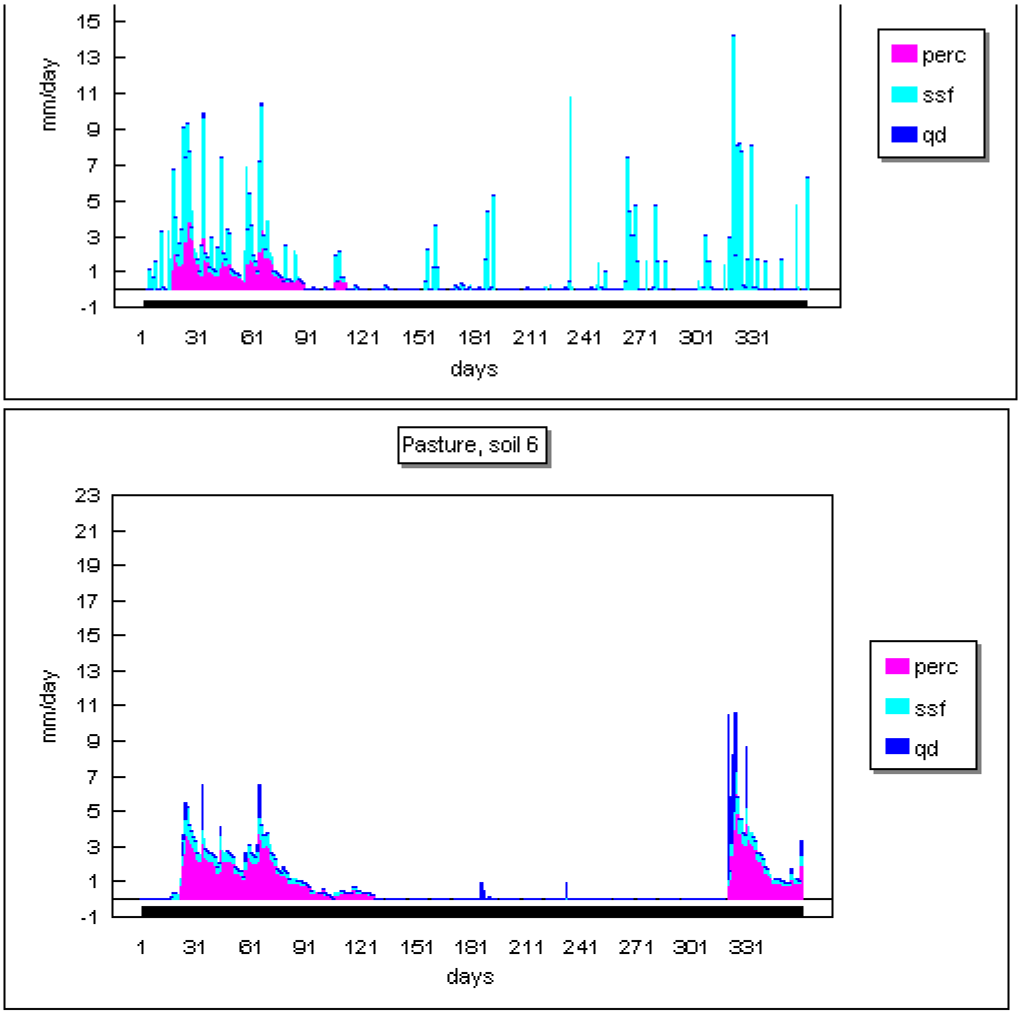


Fig. 2

The dynamics of the soil water index (soil water content / field capacity) for the same three hydrotops is shown in Fig. 3. In the hydrotop "pasture" soil is almost saturated even in summer time, while in the cropland and forest hydrotops the soil moisture is decreasing down to the wilting point in August.

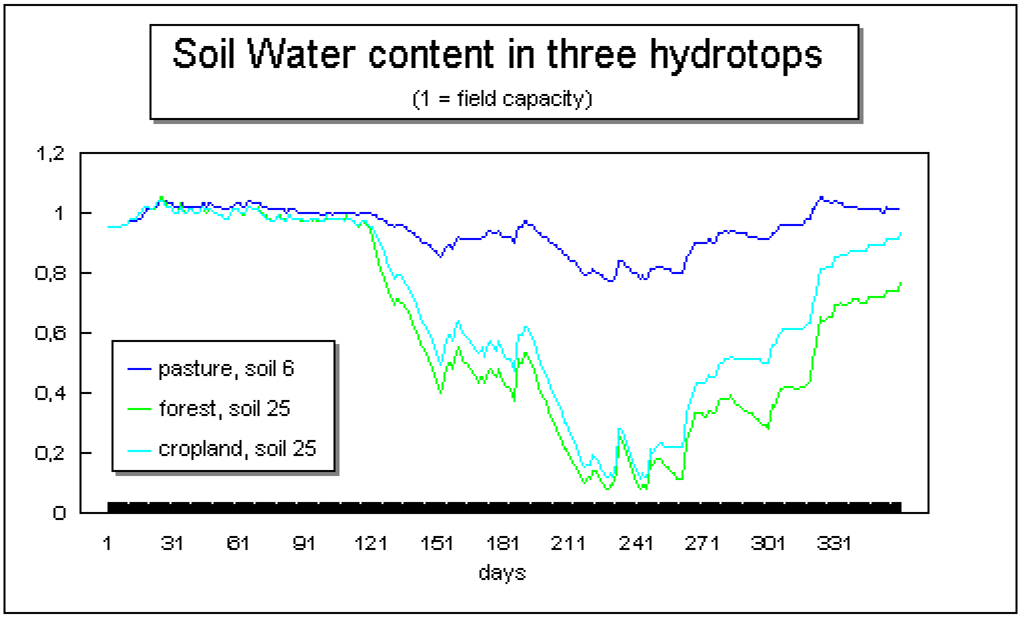


Fig. 3

Fig. 4 demonstrates the averaged water flows for subbasin 2, where the abovementioned hydrotops dominate and occupy 69% of the whole area: cropland on soil 25 covers 25%, forest on soil 25 covers 28%, pasture on soil 6 covers 16% of the subbasin area.

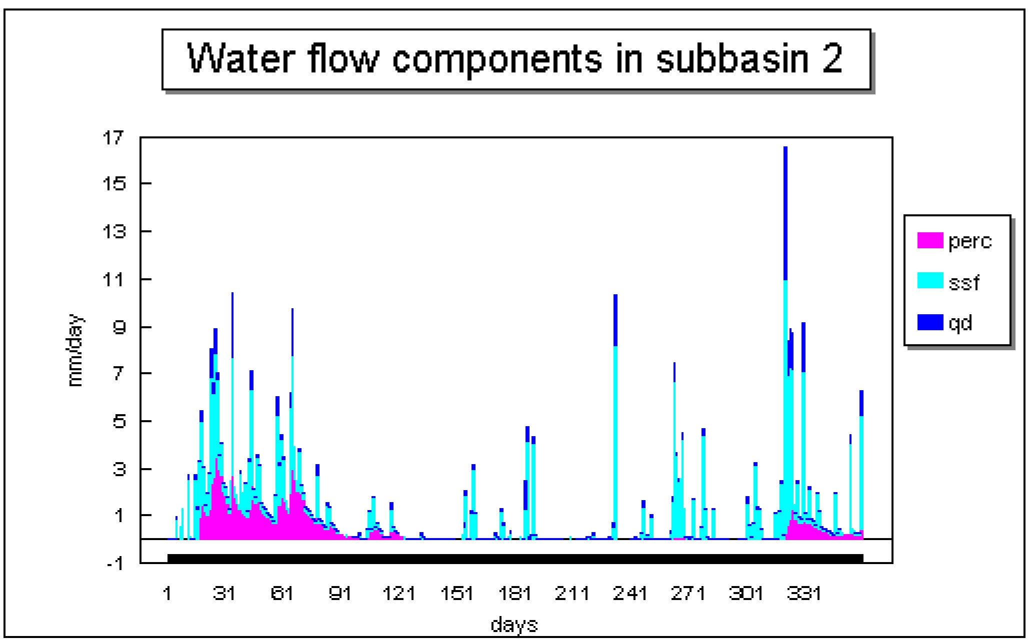


Fig. 4

Fig. 5 shows the time series of evapotranspiration and soil moisture index for the subbasin 2.

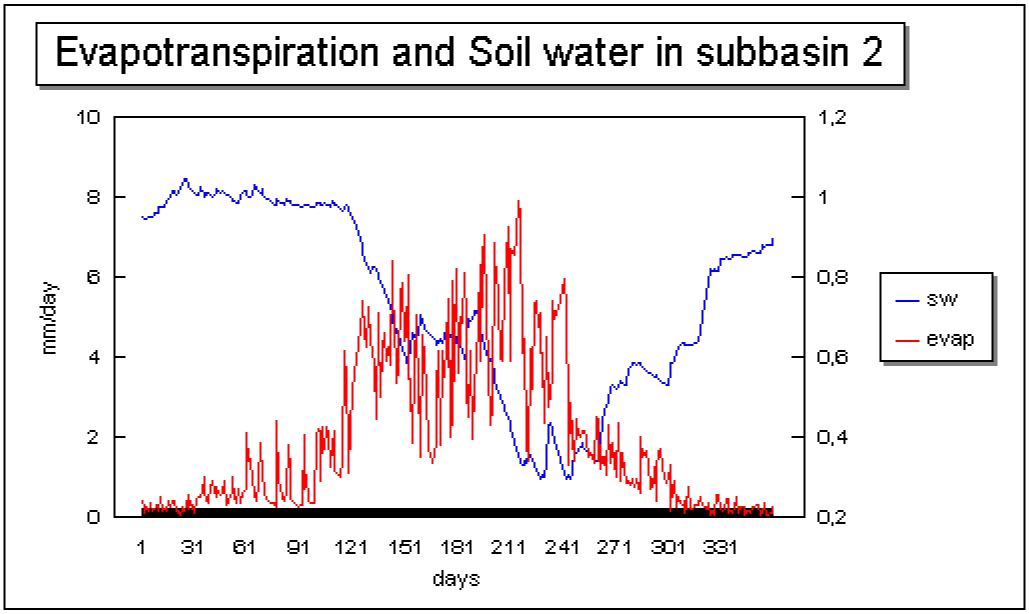


Fig. 5

The precipitation and observed versus simulated runoff in the river outlet are demonstrated in Fig. 6 for two years - 1990 and 1992.

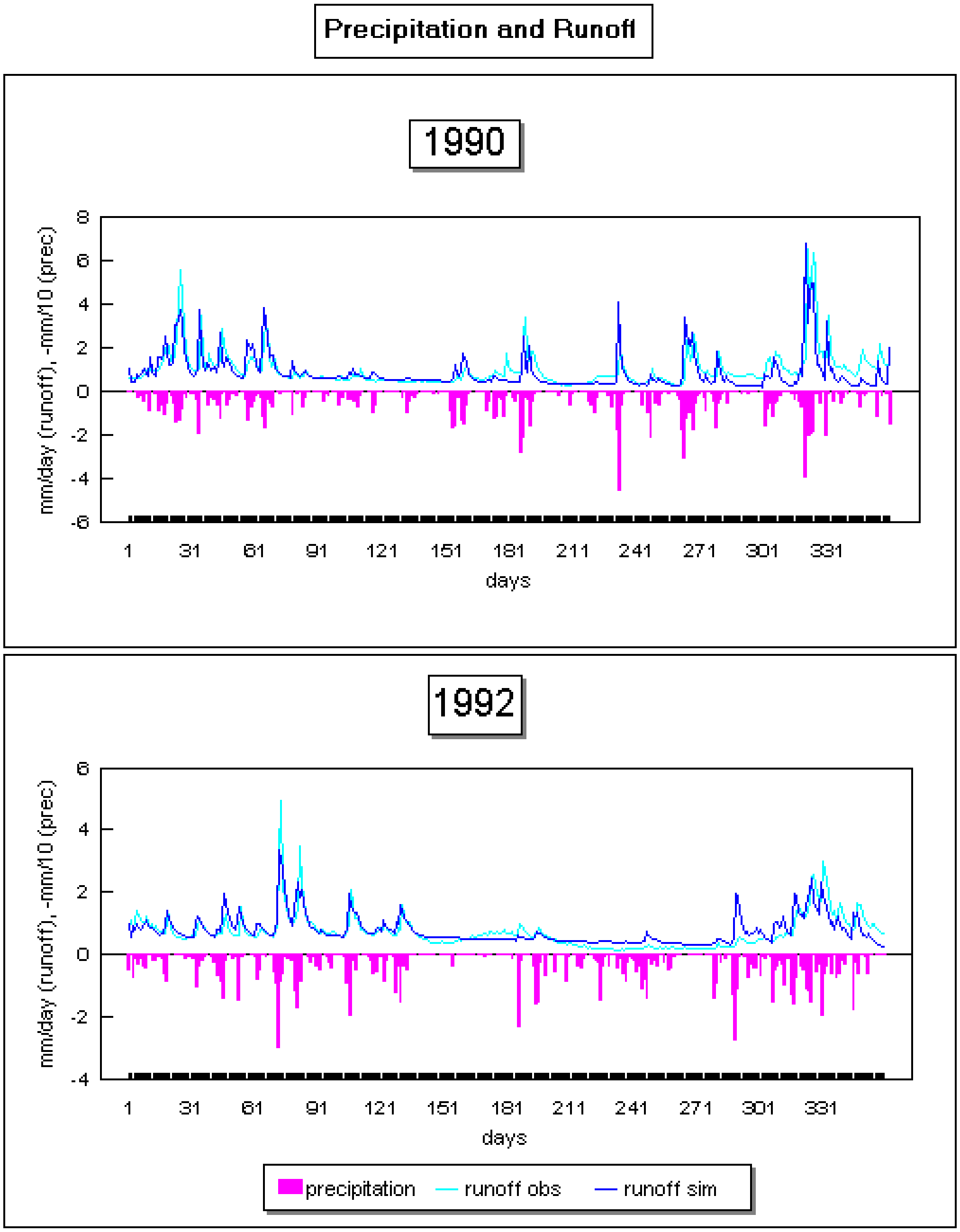


Fig. 6

The results obtained for hydrological components in Buckener Au and Dahme watersheds are satisfactory. The dynamics of water flows in hydrotops and subbasins is quite reasonable, and the Nash-Sutcliff (1979) efficiency is about 0.68 - 0.72 for these four years of simulation with daily time step. The next steps are to test the erosion, crop, and nutrient components.

After verification of results for some representative subwatersheds (Fig. 1 d), the model can be applied to other subwatersheds, and then a number of methods can be used to obtain results for the whole Elbe, like

- integration of modelling results for subwatersheds by means of a river routing model,

- nested watershed approach (upscaling based on smaller scale applications),

- application of a simplified model based on Unit Area Load estimates (which can be obtained from the modelling results).

Summary

A new version of watershed model integrating hydrology and water quality is presented in the paper. The model is based on existing tools and tries to combine their advantages and avoid overparametrization. A three-level scheme of spatial disaggregation implemented in SWIM for mesoscale basins seems to be quite reliable. Currently the model is tested in the Elbe subbasins. The intention is to extend the applicability of the model to other river basins in Europe.

Acknowledgements

We are very grateful R.Srinivasan (Blackland Research Center, Temple, TX) and J. Arnold (USDA ARS, Temple, TX) who kindly provided the SWAT codes and were always open for fruitful discussion.

The authors would like to thank Prof. Ripl (Free University of Berlin) and the Projektzentrum Ökosystemforschung, the University of the Armed Forces, Munich, and the German Weather Service (DWD) for providing data.

We are grateful to the Deutsche Forschungsgemeinschaft (DFG) and the Federal Ministry for Education, Science, Research and Technology (BMBF), which made the present work possible by providing financial support.

References

Arnold J.G., J.R.Williams, A.D.Nicks, N.B.Sammons, 1990. *SWRRB - A Basin Scale Simulation Model for Soil and Water Resources Management.* Texas A&M University Press, College Station.

Arnold J.G. 1990. ROTO - a continuous water and sediment routing model. *ASCE Proc. of the Watershed Management Symposium.* Durango, Co, p. 480-488.

Arnold J.G., Allen P.M., Bernhardt G., 1993. A comprehensive surface-groundwater flow model. *Journal of Hydrology*, 142, 47-69

Beasley D.B., L.F.Huggins, E.J.Monke, 1980. ANSWERS: A model for watershed planning. *Transactions of the ASAE*. 23(4): 938-944

Beven K.J., M.J. Kirkby, 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24, 1, 3, 43-69.

Beven K. 1995. Linking parameters across scales: subgrid parametrizations and scale dependent hydrological models. *Hydrol. Process*., 6, 279-298.

Knisel, W.G. (ed.), 1980. *CREAMS: A field scale model for chemicals, runoff, and erosion from agricultural management systems*. USDA, Conservation Research Report NO. 26, 643p.

Krysanova, V., Meiner, A., Roosaare, J., Vasilyev, A., 1989a. Simulation modelling of the coastal waters pollution from agricultural watershed. *Ecological Modelling,* 49, pp. 7-29.

Krysanova, V. and Luik, H., (eds.) 1989b. *Simulation modelling of a system watershed - river - sea bay.* Tallinn, Valgus, 428 (in Russian).

Kuchment L.S., 1992. The construction of continental scale models of the terrestrial hydrlogical cycle: ana analysis of the state-of-the-art and future prospects. J-P.O'Kane (ed.) *Advances in Theoretical Hydrology,* European Geophys. Soc. Series on Hydrol. Sciences, 1, ELSEVIER.

Meiner A., 1995. Integration of GIS and a dynamic spatially distributed model for non-point pollution management. *Proc. of the Second Int. IAWQ Spec. Conf. and Symposia on Diffuse Pollution*, Brno & Prague, Czech Republic, August 13-18, 1995.

Smith R.E..1992 *Opus, An Integrated Simulation Model for Transport of Nonpoint-Source Pollutants at the Field Scale*: Volume 1 & 2.U.S. Department of Agriculture, Agricultural Research Service, ARS-98.

Srinivasan, R., and Arnold, J.G., 1993. Basin scale water quality modelling using GIS. Proceedings, *Applications of Advanced Inform. Technologies for Manag. of Nat. Res*. June 17-19, Spokane, WA, USA.

Srinivasan, R., Arnold, J.G., Muttiah, R.S., Walker C., Dyke P.T., 1993. Hydrologic Unit Model for the United States (HUMUS). In: Sam S.Y.Wang (ed.) *Advances in Hydro-Science and -Engineering,* Vol. I.

USDA, 1992. *STATSGO - State soils geographic data base*. Soil Conservation Service, Publ. Number 1492, Washington D.C.

Williams J.R., K.G. Renard, P.T. Dyke, 1984. EPIC - a new model for assessing erosion's effect on soil productivity. *Journal of Soil and Water Conservation* 38(5): 381-383

Young R.A., C.A.Onstad, D.D.Bosch, W.P.Anderson, 1989. AGNPS: a nonpoint source pollution model for evaluating agricultural watersheds. *J.Soil and Water Cons.* 44(2): 168-173

Address for correspondence:

Valentina Krysanova, Dirk-Ingmar Müller-Wohlfeil, Alfred Becker

Potsdam Institute for Climate Impact Research
P.O.Box 601203, Telegrafenberg
14412 Potsdam, Germany
Telephone: +49-(0)331-288-2515
Fax +49-(0)331-288-2600
E-mail: valen@pik-potsdam.de

Sudhakar Mamillapalli, R. Srinivasan, J. G. Arnold, Bernard A. Engel

Effect of Spatial Variability on Basin Scale Modeling

ABSTRACT

With the integration of GIS and distributed parameter hydrologic models, a watershed can be divided into many subbasins. However, the effect of discretization on the quality of the simulated output has not been studied very widely. Using the concept of virtual basins the different soil and landuses in the subbasin can be simulated to the level of detail desired. Using a 4300km^2 watershed in Texas, the present study was undertaken to study the effect of increasing level of discretization and virtual basins, on the accuracy of the output. The results indicate that in general, increasing level of discretization and increasing the number of soil and landuse combinations simulated within each subbasin increases the accuracy of the simulation. There is a level beyond which the accuracy can't be improved, suggesting that more detailed simulation may not always be necessary. Preliminary investigations have been conducted to determine optimal configuration.

INTRODUCTION

Hydrologic models can be broadly divided into lumped parameter models and distributed parameter models. The lumped parameter approach considers the whole catchment as a single entity and maps the input rainfall excess to an output hydrograph. Though computationally efficient, this approach doesn't explicitly account for spatial variabilities present within the catchment. Chief among this type of model is the USLE(Wischmeier and Smith, 1978). Distributed models divide the catchment into a number of smaller areas (which could be square elements or subcatchments), which are assumed to be uniform with respect to the hydrologic parameters. Hydrology is simulated within each of these elements and the output routed to the outlet. Hence these models take into consideration spatial variability of the watershed. Examples of these include the AGNPS (AGricultural Non-Point Source Pollution) model (Young et al., 1987), ANSWERS (Aerial Nonpoint Source Watershed Response Simulation)(Beasely et al., 1977) and SWAT (Soil and Water Assessment Tool) (Arnold et al., 1993). Considerable time and effort are required to acquire the data, run the models and interpret resulting information. Integration with a GIS can eliminate many of these problems. Several models have been integrated with GIS which include AGNPS and GRASS GIS by Srinivasan and Engel (1994), ANSWERS and GRASS (Rewerts and Engel, 1991), and SPUR and ERDAS (Sasowsky and Gardner, 1991).

As noted before, these models either discretize the watershed into smaller elements by overlaying a square grid(ANSWERS or AGNPS) or into various subbasins (SWAT and SPUR). With the integration of these models with a GIS, it is possible to divide the watershed into a large number of elements since the GIS automatically generates the input. Hence we can consider the spatial variability to the level of detail supported by the data. However, as the number of such elements increase, so does the computation time. It is not clear from studies to date that the effect of increasing input levels of detail improves the accuracy of the simulated output. For effective use of the above tools, it is necessary to be able to discretize the watershed to an appropriate level of detail. A gross discretization may lead to poor simulation results whereas very fine discretization would require far more input data and significantly increased computation time and space (which may be important for large watersheds comprised of hundreds of subbasins) with no or little increase in accuracy. Also using the concept of virtual basins, different soil and landuse combinations within a subbasin can be simulated, instead of considering the dominant landuse and soil to be representative of the subbasin. It may not be necessary to consider all the combinations within a subbasin. The impact of increased detail in accounting for various soil and landuse combinations within a subbasin and effect of level of discretization on accuracy is studied for a watershed in Texas and presented.

OBJECTIVES

1. Quantify the effect that level of discretization has on the accuracy of output obtained from the distributed parameter hydrologic basin model SWAT.
2. Examine the impact of increased detail in simulating the soil and landuse combinations within a subbasin using the concept of virtual subbasins on the accuracy of output using SWAT.

RELEVANT LITERATURE

SWAT(Soil Water Assessment Tool), a continuous daily time step model developed by Arnold et al.(1993), was obtained by adding a new routing structure to the SWRRB model(Arnold et al., 1990, Williams et al.,1985) so as to remove the restriction of only being able to simulate 10 subwatersheds in the case of SWRRB. The new routing structure of SWAT routes and adds flows down through the basin reaches and reservoirs. Apart from this, changes were incorporated to simulate lateral flow, ground water flow, reach routing transmission losses, and sediment and chemical movement through ponds, reservoirs, streams and valleys. SWAT is capable of simulating hundreds of subwatersheds for periods of 100 years or more. The major components of the model include hydrology, weather, sedimentation, soil temperature, crop growth, nutrients, pesticides, ground water and lateral flow, and agriculture management. Additional details about the model can be found in Arnold et al. (1994).

SWAT allows for considerable flexibility in watershed discretization. The watershed can be divided into cells and/or subwatersheds. Different parts of the watershed can be divided differently. The dominant soil and landuse within each subbasin is considered to be the soil and landuse of the subbasin. However, in order to account for multiple soil and landuse combinations, the concept of virtual subbasins was incorporated into

SWAT. Instead of assuming the dominant soil and landuse to be the soil or landuse of the subbasin, each subbasin is discretized into virtual areas (referred to as virtual basins), each having a unique soil and landuse combination without reference to their spatial positioning within the subbasin. This is similar to the concept of Hydrologic Response Units (HRU's) given by Maidment (1991). The hydrologic response is generated within each of these virtual areas and then the weighted average (by area) of the response from these virtual subbasins is taken to be the output of the subbasin. Since there can be large numbers of such combinations, a threshold is set. Only soil and landuse combinations forming a proportion larger than that of the threshold are considered. The threshold is arbitrary and is set by the user.

The way the dominant soil and landuse are chosen in SWAT is that, first the dominant landuse is determined within the subbasin, and within this landuse the dominant soil prevalent is determined. This soil and landuse are taken to be representative of the subbasin. It should be noted this may not be the same as the dominant soil and landuse combination. For example this may occur within a subbasin where there are two different landuses A and B each occupying say 60% and 40% and landuse A has 10 different soils occupying 10% of the area while landuse B has just one soil occupying the whole area. In this case the soil landuse combination chosen will be the A landuse with one of the soils even though the other combination may occupy a larger proportion of the area. Since between soil and landuse, landuse affects stream flow more, this approach seems logical. In the case of the virtual basin approach, based on the threshold on landuse set by the user all the landuses occupying an area greater than the threshold are selected. Within these landuses, soils forming a proportion greater than that of the soil threshold set by the user is used to select the soils. Both the thresholds are set by the user. The effect of varying the threshold on the output obtained is studied.

A GRASS GIS interface to the SWAT model was developed by Srinivasan and Arnold (1994). Given the appropriate data layers and data bases the interface extracts the data and writes the SWAT input file. A number of tools were incorporated allowing for automatic extraction of various inputs. These include tools for accessing the appropriate databases, hydrologic tools to retrieve the topographic and other attributes including automatic generation of the routing structure and aggregarion tools to aggregate the inputs at the subwatersed level. Once the input files are generated the model can be run and then the results visualized using the output interface. The GIS interface, facilitated to a large extent the study conducted.

Among studies made to determine the impact of level of discretization on output of basin scale models are that of Wood et al. (1988) and that of Sasowsky and Gardner (1991). Wood et al. introduced the concept of the representative elementary area (REA) in hydrologic modeling. They refer to REA as a fundamental building block of catchment modeling. They argue that at smaller scales, actual patterns of variability of topography, soil or rainfall lead to differences in the output even though the underlying distribution is the same. As larger and larger scales are considered, more and more of the variability is sampled and then finally an area is obtained whose hydrologic response can be considered to be the net effect of the individual point hydrologic responses within the subbasin or basin. So a basin with all its variation in soils, topography, and weather can be represented by these REAs without much loss in quality of the output. To prove the existence of the REA, Wood et al. discretized the Coweta River experimental catchment in North Carolina which had an area of 17km^2, into 3, 19, 39 and 89 subcatchments by the method described by Band and Wood (1986). In order to emulate point hydrologic response which can be then averaged to form the basin hydrologic response, they applied TOPMODEL (Beven and Kirby, 1979; Beven, 1986) within each 30m pixel comprising the catchment. Then pixel output was aggregated to form the subbasin response. The subbasin responses where then arranged in increasing order of their areas and a running average of 15 subcatchments, moving in steps of 5, was taken. The mean area within each window was plotted against the mean average response. The graphs indicated that the areal response stablized at around 1km^2 area. The size was the same for all the outputs studied. Thus they concluded the REA for this catchment was 1km^2. They made further studies and remarked that the size of the REA is governed primarily by the topography. Soil and rainfall variability didn't have a big role in determining the size of the REA. In this study, variability in only soil, rainfall and topography were studied. In general, large catchments in addition to the above have land use variability to consider.

Sasowsky and Gardner (1991) used three different configurations of the 146km$^2 Walnut Gulch watershed in Arizona: >= 2nd, >= 4th, >= 13th stream order with 28, 15, and 1 channel segments and 66, 37, and 3 contributing areas and made SPUR runs for each of these configurations. The runoff at the outlet was compared with observed data and then the results seemed to imply the >= 4th order stream network gave as good as results as the >= 2nd and hence they concluded that REA exists for the basin under consideration. However, again this study doesn't consider land use variability and they admit the evaluation criteria used affect the conclusions that could be drawn. Also calibration (the curve number) needed to be done on the model and again the the model evaluation results depended on the changes made in the curve number. Due to ambiguities, it is difficult to conclude that the study establishes the existence of the REA.

METHDOLOGY

A watershed in Texas of size 4297 km^2 was used in this study. It has originally been subdivided into 40 subbasins, composed of agriculture and range land. Figure 1 shows the watershed with the 40 subbasins. Using the "r.watershed" tool within the GRASS GIS and the 1:250,000 DEM, the watershed was discretized into 4, 8, 14, 20, 24, 29, 35, 40 and 54 subbasins. Measured stream flow data was available at two locations within the watershed. Since both these gages are not located at the outlet of the subbasin, the simulated flow draining into the basin where the gage was located was extracted and compared with the output. Statistics used in the comparison are the coefficient of determination, and the coefficient of efficiency of Nash and Sutcliffe (1971). A coefficient of efficiency of 1 indicates perfect agreement. If the results are highly correlated but biased, then the coefficient of efficiency will be less than the coefficient of determination (Aitken, 1973).

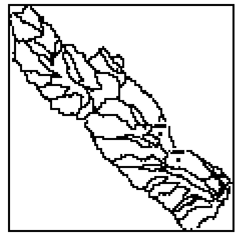


Figure 1: The Bosque Watershed

Simulations were made both for the dominant case where the dominant soil and landuse within the basin was considered to be the soil and landuse of each subbasin and the virtual basin approach with thresholds ranging from 5% to 20% for landuse and 10% to 40% for soil. For example a threshold of 10% for soil and 5% for landuse indicates that landuses which form at least 5% of the subbasin area and soils which form at least 10% of the area within each of the selected landuses will be taken as virtual basins. Results for all these cases are presented here. In the table of results the different thresholds are mentioned as "landuse threshold and soil threshold". For example "10% and 20%" in the results table indicate that 10% is the landuse threshold and 20% the soil threshold used for the corresponding results. Observed flow data was available for years 1965 to 1974 and 1975 to 1984 for two different USGS gages 5000 and 5200 within the basin. Simulations were made for these time periods and the results compared. A single simulation was not done for both these time periods since the rain gage data available changed in 1975. No calibration whatsoever was attempted throughout the study, so that the impact of spatial variability alone can be studied. Also to remove the impact of weather variability, the Thiessen polygon average of all the weather gages present in the basin was taken to be the rainfall and temperature data for all the subbasins within the basin.

RESULTS

Statistics, including mean, standard deviation, coefficient of efficiency, and coefficient of determination were computed for the simulations described previously. Results are presented for both USGS gages 5000 and 5200 within the study basin for years 1965-1974 and 1975 to 1984. These gages are upstream of the outlet, so the number of basins draining into the basin having these gages is different from the total number of basins in the watershed. In the result tables, the number of basins draining into these gages, as well as the total number of basins in the watershed are presented. The coefficient of efficiency (COE) is used as a measure of accuracy of the simulated results and presented here, though other measures generally followed the same trend.

Table 1: Results Obtained when Runoff Simulated using Various Approaches for Different Basin Configurations for Years 1965 to 1974 for Gauge 5000

No of Basins		Coefficient of Efficiency				
Watershed	At Gage	Dominant	20 and 40*	15 and 30*	10 and 20*	5 and 10*
54	28	0.72	0.72	0.73	0.74	0.71
40	21	0.72	0.72	0.73	0.73	0.72
35	20	0.69	0.69	0.68	0.73	0.72
29	18	0.70	0.70	0.71	0.73	0.71
24	17	0.70	0.70	0.71	0.73	0.71
20	12	0.70	0.70	0.71	0.73	0.71
14	6	0.69	0.69	0.68	0.69	0.70
8	4	0.58	0.58	0.60	0.68	0.68
6	2	0.35	0.35	0.34	0.46	0.69
4	1	0.31	0.31	0.31	0.38	0.68

*Landuse and Soil Thresholds

Table 2: Results Obtained when Runoff Simulated using Various Approaches for Different Basin Configurations for Years 1965 to 1974 for Gauge 5200

No of Basins		Coefficient of Efficiency				
Watershed	At Gage	Dominant	20 and 40*	15 and 30*	10 and 20*	5 and 10*
54	35	0.73	0.74	0.74	0.74	0.73
40	25	0.72	0.72	0.73	0.73	0.74
35	22	0.69	0.69	0.69	0.73	0.74
29	20	0.70	0.70	0.71	0.73	0.74
24	19	0.71	0.70	0.71	0.73	0.74
20	14	0.71	0.71	0.71	0.73	0.74
14	8	0.68	0.68	0.68	0.69	0.73
8	6	0.60	0.62	0.61	0.68	0.72
6	5	0.43	0.45	0.43	0.52	0.73

*Landuse and Soil Thresholds

Table 1 gives the coefficient of efficiencies obtained for different basin configurations using dominant soil and landuse approach and various soil and landuse thresholds for the virtual basin approach for years 1965-1974 for gage 5000. As noted before, no calibration was attempted whatsoever of the results. From the table it is clear that as the number of basins increased so does the coefficient of efficiency. For example when using the dominant soil and landuse approach using just 1 basin, the coefficient of efficiency is 0.31 while with 28 basins the coefficient of efficiency rose to 0.72. Within the same configuration, accuracy in simulation increases as more and more soil and landuse combinations are simulated chosing smaller soil and landuse thresholds, with the best results obtained with the 5% landuse and 10% soil threshold. However, the increase in accuracy is minimal as more and more basins are chosen for simulation, i.e. for more and more detailed configurations. For example with 28 basins there is practically no increase in accuracy between the dominant approach and the other configurations with all giving a COE of around 0.72. However, with 1 basin using the dominant approach, the COE is 0.31 which increases to 0.68 with landuse and soil thresholds of 5% and 10%. The results for gage 5200 given in Table 2 are very similar.

Results for 1975 to 1984 are given in Tables 3 and 4 for gages 5000 and 5200 respectively. The results are relatively poor compared to the previous case (with the maximum COE achieved at 0.74 for gages 5000 and 5200 for 1965-1974 but 0.48 and 0.49 for years 1975-84 for gages 5000 and 5200 respectively). The conclusions drawn however are the same as in previous case. As the number of basins used in simulations increased, so did the accuracy, for each soil and landuse threshold chosen. Also within each configuration as the number of soil and landuse combinations chosen increased, so did the accuracy. However, this increase was minimal for detailed configurations compared to less detailed configurations. From these results, two conclusions can be drawn. Firstly, the increase in accuracy can be obtained either by increasing the number of basins used in simulation or by increasing the number of soil and landuse combinations within each subbasin. Secondly there is a limit in the accuracy that could be obtained. Increased detail in soil and landuse combinations simulated or the basin configuration may not give rise to better results, but on the other hand increases the number of simulations made~(since simulations are needed for each soil, landuse and basin combination selected. For example if the configuration has 40 subbasins and within each subbasin on the average 5 soil and landuse combinations have been selected a total of 200 combinations need to be simulated). For example considering Table 1, using 54 subbasins (which leads to 35 subbasins flowing into the stream gage), the number of combinations simulated in the dominant case was 54, while using a 5% threshold for soil and 10% threshold for landuse the number of combinations were 362, however there is no increase in COE (both are 0.73). Similarly, examining the 5% and 10% threshold column in the table, there is practically no increase in accuracy with either 5 subbasin or 35 subbasins. However, this doesn't indicate that using 5 subbasins with the above threshold will always give the best results. For example for gage 5200 for year 1975 to 1984 (Table 4), the COE increased from 0.43 for 5 subbasins to 0.49 for 35 subbasins. Even though the increases is not much, it shows that 5 basins will not always give as good results as that of 54 subbasins.

Table 3: Results Obtained when Runoff Simulated using Various Approaches for Different Basin Configurations for Years 1975 to 1984 for Gauge 5000

No of Basins		Coefficient of Efficiency				
Watershed	At Gage	Dominant	20 and 40[a]	15 and 30[a]	10 and 20[a]	5 and 10[a]
54	28	0.33	0.37	0.39	0.42	0.48
40	21	0.30	0.33	0.34	0.38	0.45
35	20	0.32	0.34	0.35	0.41	0.47
29	18	0.34	0.36	0.37	0.41	0.47
24	17	0.34	0.36	0.38	0.40	0.47
20	12	0.35	0.36	0.39	0.42	0.47
14	6	0.25	0.26	0.23	0.33	0.45
8	4	0.16	0.13	0.14	0.32	0.39
6	2	-0.07	-0.06	-0.04	0.14	0.38
4	1	-0.06	-0.10	-0.10	0.09	0.39

[a]Landuse and Soil Thresholds

Table 4: Results Obtained when Runoff Simulated using Various Approaches for Different Basin Configurations for Years 1975 to 1984 for Gauge 5200

No of Basins		Coefficient of Efficiency				
Watershed	At Gage	Dominant	20 and 40[a]	15 and 30[a]	10 and 20[a]	5 and 10[a]
54	35	0.39	0.42	0.44	0.47	0.49
40	25	0.35	0.37	0.38	0.43	0.48
35	22	0.35	0.37	0.39	0.45	0.49
29	20	0.38	0.39	0.41	0.45	0.49
24	19	0.37	0.39	0.42	0.44	0.49
20	14	0.38	0.38	0.42	0.45	0.49
14	8	0.30	0.30	0.29	0.37	0.48
8	6	0.22	0.19	0.21	0.36	0.44
6	5	0.04	0.04	0.06	0.21	0.43

[a]Landuse and Soil Thresholds

Even though the conclusions that could be drawn are similar for both 1965-74 and 1975-84, there seems to be a vast difference in accuracy. This may be because 1975-84 is drier (annual average 783.1 mm) compared to 1965-74 (annual average 859.1 mm) and SWAT is known to do better in wet conditions. Also the landuse might have changed between the two different simulations. This might have lead to relatively poor results. This however leads to interesting conclusions. For example looking at Table 2 one might conclude that reasonable results would be obtained using dominant soil and landuse with 14 basins with no increase in accuracy going to more basins or more detailed soil and landuse combinations. However, from Table 4, it is clear that this may not lead to the best possible results. The conditions being simulated seem to have an impact on the optimum basin configuration.

Next, some studies were made to determine if possible, the optimal configuration. There are two different aspects to this problem. First is to chose the appropriate landuse and soil thresholds given a particular basin configuration and the second is to chose a appropriate basin configuration. The first of these two problems is addressed here.

For each of the basin configurations and soil and landuse thresholds, the curve number was plotted against the percentage of area with the corresponding curve number. Also plotted on this curve is the curve number distribution obtained choosing this configuration but with 0 soil and landuse threshold. This is done for each of the basin configurations and the various thresholds for which simulation runs were made, in each the comparision being made with the corresponding curve number distribution of the same configuration except 0 soil and landuse threshold. In general it was found that, the closer the distribution is to this more detailed, 0 soil and landuse threshold distribution the better were the results. In order to quantify the results, the coefficient of efficiency was calculated between the detailed distribution mentioned above to that of different soil and landuse thresholds for the same configuration and presented in Table 5.

A higher COE between the distributions in general lead to higher COE between the observed and simulation results and vice versa. For example when the watershed is comprised of 4 subbasins, the COE is very low for all soil and landuse thresholds except that when 5% and 10% thresholds are used. Looking at the results in Tables 1 and 3, the COE's are low for all combinations except for the 5% and 10% thresholds. A low COE in Table 5 doesn't necessarily indicate poor results in all cases. For example for 54 subbasins, even though COE increased from 0.60 in the dominant case to 0.93 for the 5% and 10% thresholds case, the accuracy of the simulations were identical for years 1965 to 1974 (Tables 1 and 2), with the COE between observed and simulated results being above 0.70. However, for 1975 to 1984 there is an improvement in the simulation results for this same configuration. For example in Table 3 the COE increased from 0.33 in the dominant case to 0.48 for the 5% and 10% thresholds. In most cases there is a big jump in COE from all other threshold to the one with 5% and 10% soil and landuse thresholds. Also, corresponding to this there is a jumb in COE, whih is especially noticeable in 1975-1984 results. Checking for the COE between the actual and the simulated distributions may be a convenient way of determining the optimal thresholds within a particular configuration.

Table 5. Relation between Simulated and Actual Soil and Landuse Distributions

Number of Basins	Coefficient of Efficiency				
	Dominant	20 and 40[a]	15 and 30[a]	10 and 20[a]	5 and 10[a]
54	0.60	0.55	0.63	0.66	0.93
40	0.66	0.68	0.70	0.75	0.93
35	0.63	0.65	0.71	0.69	0.91
29	0.65	0.64	0.70	0.70	0.91
24	0.40	0.61	0.64	0.63	0.90
20	0.40	0.59	0.66	0.62	0.89
14	0.41	0.57	0.63	0.69	0.88
8	0.09	0.59	0.62	0.65	0.84
6	0.22	0.54	0.56	0.70	0.81
4	0.14	0.36	0.36	0.58	0.81

[a]Landuse and Soil Thresholds

CONCLUSIONS

With the integration of GIS and distributed parameter hydrologic models, a watershed can be divided into many subbasins. However, the effect of discretization on the quality of the simulated output has not been studied very widely. Using the concept of virtual basins, the different soil and landuses in the subbasin can be simulated to the level of detail intended. Using a 4300km^2 watershed in Texas, the present study was undertaken to examine the effect of increasing level of discretization and virtual basins on the accuracy of the output. The basin was divided into numerous configurations using the "r.watershed" tool within the GRASS GIS. Simulations were made for the various configurations and within each of these configurations various soil and landuse thresholds for selection of virtual basins. The results indicate that in general, increasing level of discretization and increase in the number of soil and landuse combinations simulated within each subbasin increases the accuracy of the simulation. There is a level beyond which the accuracy can't be improved, suggesting that more detailed simulation may not always lead to better results. It is important to determine the optimal configuration, so that reasonable results could be obtained without the necessity of detailed simulations. It was also noted that from the different time periods considered, some of the coarser levels of discretization may perform well for one period, but not perform well for another period, whereas the finer simulations performed well throughout.

In order to examine the results further, the proportion of the watershed area having different curve numbers is plotted for each of the soil and landuse thresholds against that obtained when all soil and landuse combinations are considered within that particular configuration. The coefficient of efficiency is calculated and it was seen that in general within a particular configuration as smaller and smaller thresholds are considered, the curve number distribution better matches that using 0 thresholds and the simulation results improved. This approach seems to hold promise to determine the optimal soil and landuse threshold that need to be chosen within a particular configuration, though the optimal basin configuration can not be determined.

There are some limitations to this study, one of them being the fact that for other model outputs like sediment, apart from curve number, other soil and topographic properties play a major role in determining the output. So in such cases, matching the curve number distribution may not indicate better results. Also no method to determine the optimal configuration has been given. Also the effect of weather variability has not been considered with a single weather used to simulate all the configurations. Using a detailed configuration the weather might be better represented as compared to coarser configurations. More studies need to be conducted on other watersheds of different sizes, and variability in order to validate these results.

REFERENCES

Assessing systematic errors in rainfall-runoff models. Journal of Hydrology, 20: 121--136.

Arnold, J.G., Engel, B.A., and Srinivasan, R. 1993. Continuous time grid cell watershed model. In Heatwole, C.D. (Ed.), Application of Advanced Information Technologies: Effective Management of Natural Resources, 2950 Niles Rd, St. Joseph, Michigan 49085-9659 USA. Information and Electrical Technologies Division of ASAE, American Society of Agricultural Engineers.

Arnold, J.G., Williams, J.R., Nicks, A., and Sammons, N.B. 1990. SWWRB-A basin Scale Simulation Model. College Station: Texas A&M Press.

Arnold, J.G., Williams, J.R., Srinivasan, R., King, K.W., and Griggs, R.H. 1994. SWAT-Soil Water Assesssment Tool. 808 East Blackland Rd, Temple, TX-76502: USDA , Agricultural Research Service and Grassland, Soil and Water Research Laboratory.

Band, L. E, 1986. Topographic partition of watershed with digital elevation models. Water Resources Research, 22 (1): 15--24.

Beasely, D.B., Huggins, L. F., Monke, E.J. 1980. ANSWERS: A model for watershed planning. Transactions of the ASAE, 23(4): 938-944.

Runoff production and flood frequeny in catchments of order n: an alternative approach. In Gupta, V.K., Rodriguez-Itrube, I., and Wood, E.F. (Eds.), Scale Problems in Hydrology.

Beven, K.J. and Kirkby, M.J. 1979. A physically based variable contributing area model of basin hydrology. Hydrological Sciences Bulletin, 24(1): 43--69.

Engel, B.A., Srinivasan, R., and Rewerts, C. 1993. A spatial decision support system for modeling and managing agricultural nonpoint source pollution. In Goodchild, M. F., Parks, B. O., and Steyart, L. T.(Eds.), Environmental Modeling with GIS, (pp. 231-237). Oxford University Press, New York, NY.

Maidment, D.R. 1991. GIS and hydrologic modeling. In Prepared for Presentation at the First International Symposium/Workshop on GIS and Environmental Modeling, Boulder, Colorado.

Nash, J.E. and Sutcliffe, J.V. 1971. River flow forecasting through conceptual models. Journal of Hydrology, 13: 297--324.

Rewerts, C.C. and Engel, B.A. 1991. ANSWERS on GRASS: Integrating a watershed simulation with GIS. Number ASAE Paper No. 91-2621. American Society of Agricultural Engineers, St.Joseph, MI.

Sasowsky, K.C. and Gardner, T.W. 1991. Watershed configuration and geographic information system parametrization for SPUR model for hydrologic simulations. Water Resources Bulletin, 27(1): 7--18.

Srinivasan, R. and Arnold, J.G. 1994. Integration of a basin scale water quality model with gis. Water Resources Bulletin, 30(3): 453--462.

Williams, J.R., Nicks, A.D., and Arnold, J.G. 1985. SWWRB, a simulator for water resources in rural basins. ASCE Hydraulics Journal, 111(6): 970--986.

Wischmeier, W.M. and Smith, D.D. 1978. Predicting rainfall erosion losses - a guide to conservation planning. Technical Report Agri. Handbook No. 537, Science and Education Administration, USDA.

Wood, E.F., Sivapalan, M., Beven, K., and Band, L. 1988. Effects of spatial variability and scale with implications to hydrologic modeling. Journal of Hydrology, 102: 29--47.

Young, R.A., Onstad, C.A., Bosch, D.D., and Anderson, W.P. 1987. AGNPS, agricultural non-point-source pollution model: A watershed analysis tool. Technical Report Report 35, U.S Department of Agriculture.

MODELING WISTER LAKE WATERSHED USING A GIS-LINKED BASIN-SCALE HYDROLOGIC/WATER QUALITY MODEL

Tharacad S. Ramanarayanan, Raghavan Srinivasan, and Jeffrey G. Arnold

ABSTRACT

Basin scale hydrologic/water quality models that are capable of predicting long-term effects of land management are necessary to aid water quality assessment of river systems. SWAT is a GIS-linked basin scale model capable of simulating hydrology and water quality for 100 or more years. We incorporated the instream kinetics of an instream water quality model, QUAL2E into SWAT, thus improving the overall capability of the SWAT model. The combined model was applied to simulate hydrology and water quality in Wister Lake watershed situated in Arkansas River basin. The results from the `first-cut' preliminary analysis are presented in this paper. The model predicted stream flow, water temperature and dissolved oxygen satisfactorily, but needs improvement in nutrient simulation capability.

INTRODUCTION

River systems are a major source of water for agricultural and urban water needs. Water quality assessments of river systems are becoming critical throughout the country and there is a real concern about the sustainable supply of quality water and the health of the water bodies. River systems should be continuously monitored to assess the effects of different land management practices on water quality. But, long-term continuous monitoring is not currently being conducted due to high costs. Therefore, there is a need for an alternate tool such as a basin-scale hydrologic/water quality model that is capable of predicting the effects of land management with reasonable level of accuracy.

SWAT (Arnold *et al*., 1993) is a basin-scale hydrologic/water quality model developed to predict the effects of alternative river basin land use management decisions on water, sediment, and chemical yields. SWAT operates on a daily time step and is capable of simulating 100 or more years. Major components of the model include hydrology, weather, erosion, soil temperature, crop growth, nutrients, pesticides, subsurface flow, and agricultural management. SWAT offers distributed parameter and continuous time simulation with flexible watershed configuration, automatic irrigation and fertilization, inter-basin water transfer, and lake water quality simulation capabilities.

Until now the instream nutrient dynamics were not considered in the SWAT model. In order to simulate the instream dynamics, the kinetic routines from an instream water quality model, QUAL2E (Brown and Barnwell, 1987), were modified and incorporated in SWAT. In this paper we have described the instream water quality component of SWAT and presented the preliminary results from the application of the model to Wister Lake watershed situated in eastern Oklahoma and western Arkansas.

Since recent past, GIS has been playing an important role in natural resources modeling and proving to be an effective tool for non-point source (NPS) pollution models (Pelletier, 1985; Hession and Shanholtz, 1988; Srinivasan and Arnold, 1994). A continuous time, distributed parameter model like SWAT overcomes some of the limitations of single-event models (Arnold *et al.*, 1995). SWAT considers a basin or watershed divided into subbasins based on topography, soil, and land use and thus preserves the spatially-distributed parameters of the entire basin and homogeneous characteristics within a subbasin. But manual collection of inputs for such models is often difficult and tedious due to the level of aggregation and the nature of spatial distribution. For this a GIS has been proven to be an excellent tool to aggregate and organize input data for distributed parameter hydrologic/water quality models (Tim *et al.*, 1991; Rewerts and Engel, 1991; Srinivasan *et al.*, 1993; Rosenthal *et al.*, 1995).

The SWAT model has been integrated with a raster-based GIS, GRASS (Shapiro *et al.*, 1992), designed and developed by the Environmental Division of the U.S. Army Construction Engineering Research Laboratory. The SWAT/GRASS interface (Srinivasan and Arnold, 1994) consists of three modules: (a) project manager, (b) input extractor and aggregator, and (c) input editor. The project manager interacts with the user to collect, prepare, edit, and store the basin and subbasin information to be formatted into SWAT input files. The input extractor and aggregator uses a variety of hydrologic tools (Srinivasan and Arnold, 1993) to derive SWAT input information from GRASS raster/site map layers such as basin boundary map with subbasin delineation, digital elevation map (DEM), soils map, land use/land cover map and weather generator/station location map. In addition the reservoirs, inflow, pond and lake data are collected directly from the user. The input editor is used to either view, edit or check the data collected from the previous phase, which are arranged in different data forms. Rosenthal *et al.* (1995) used this interface to aggregate SWAT input data for the Lower Colorado River basin of Texas and found that the SWAT/GRASS interface reduced the data collection and manipulation time by several folds, and allowed the user to modify and analyze various alternative management practices rather easily. Further details about the interface are given by Srinivasan and Arnold (1994).

INSTREAM WATER QUALITY COMPONENT

The water quality parameters simulated by the instream water quality component are algae as Chlorophyll-A (Chl-a), dissolved oxygen (DO), carbonaceous biochemical oxygen demand (CBOD), organic nitrogen (OrgN), ammonium nitrogen (NH4-N), nitrite nitrogen (NO2-N), nitrate nitrogen (NO3-N), organic phosphorus as P (OrgP), and soluble phosphorus as P (SolP). The basic mass transport equation used in QUAL2E is given by

$$\frac{\partial C}{\partial t} = \{dispersion\} + \{advection\} + \{reaction\} + \{source / sink\} \quad \text{... 1}$$

and the general first-order rate equation for every constituent is given by:

$$\frac{\partial C}{\partial t} = r_i \, C_i^{n+1} + p_i \quad \text{...2}$$

where C is the constituent concentration (ML-3), t is time (T), \`i' is the spatial element, \`n' is the time segment, \`r' is the first-order rate constant (T-1), and \`p' is the internal constituent source/sink (ML-3T-1). In QUAL2E each reach is divided into several computational elements and for each time step the resultant differential equation is solved numerically by implicit backward difference technique. In the instream water quality component of SWAT, the diffusion is ignored assuming complete mixing within a reach. The advection component is considered in the SWAT's existing routing component. Thus, in the instream water quality component only the first-order rate decay is considered, which depends on the travel time of the constituent within a reach by considering a reach as a single computational element. First-order decay relationships for Chl-A, nutrients, CBOD, and DO used in QUAL2E were adopted in SWAT with necessary modifications.

Temperature

In the instream water quality component of SWAT, water temperature simulation is not physically based, but is estimated from the air temperature based on a relationship developed by Stefan and Preud'homme (1993) through regression analysis of many river observations. The relationship is given by

$$T_w = 5.0 + 0.75\,T_a \quad \text{... 3}$$

where Tw and Ta are temperature of the water and air (C), respectively. It can be seen from the above equation that the water will be warmer than air if the air temperature is below 20C, which may be consistent with most rivers. Exceptions to this will include small streams under heavy shading, places where the stream temperature is influenced by anthropological activity (e.g. discharge of waste heat from a power generation plant).

Temperature Effects on Rate Coefficients

The rate coefficients (ri) depend on the water temperature. The temperature correction for the rate coefficients is given by:

$$X_T = X_{20}\,\theta^{(T-20)} \quad \text{... 4}$$

where XT is the temperature-corrected coefficient at the local water temperature T, X20 is the value of the coefficient at 20C, and is an empirical constant for each reaction coefficient. The default values for given by Brown and Barnwell (1987) are used in SWAT. Every temperature dependent first-order reaction rate coefficient is corrected using the above equation.

LOADING FUNCTIONS

Prior to the addition of the instream kinetics , SWAT did not predict Chl-A, CBOD, and DO loads from sub-basins to the streams. Appropriate loading estimations for these were established from the available literature. Our aim is to define the loading of these variables as a function of flow, nitrogen, phosphorous, organic matter, and temperature, whose dynamics are already defined in SWAT. A brief description of these functions are given below.

Algae/Chlorophyll-A

Chl-A is assumed to be directly proportional to suspended algal biomass concentration in water. Therefore the algal biomass loading to the stream is estimated as Chl-A loading from the sub-basins. We used the relationship developed by Cluis et al. (1988). They developed relationships between nutrient enrichment index (TN:TP), Chl-a and algal growth potential in the North Yamaska river, Canada. This lead to the functional relationship of the form:

$$(AGP + Chl\_A)Q = a\left(\frac{TN}{TP}\right)^b \quad \text{... 5}$$

where AGP is the algal growth potential (mg/l), Chl_A is Chl-A (g/l), Q is flow (m3/s), TN is the total kjeldhal nitrogen (TKN = OrgN + NH4-N) load (kg) and TP is total phosphorous load (kg of P), and \`a' and \`b' are coefficient and exponent, respectively. Cluis et al. (1988) presented the values of \`a' and \`b' for different seasons, and in SWAT the summer values (a = 7.25 and b = -4.68) were used. In addition, through their analysis they established that 1 mg/l of AGP is equivalent to 1g/l of Chl-A.

This relationship is based on regression analyses and not physically based. We resolved to using this type of expression because, (a) the actual physical relationships are very complex, and (b) studies conducted to define these actual physical relationships are very limited.

Carbonaceous Biochemical Oxygen Demand (CBOD)

The SWAT loading function for the ultimate CBOD is based on a relationship given by Thomann and Mueller (1987):

$$L_u = \frac{2.3 C_{org}}{q} \quad \text{... 6}$$

where Lu is ultimate CBOD (mg/l), Corg is organic carbon load (mg), and q is the flow (*l*). The organic carbon concentration is calculated as

$$C_{org} = OC_1 * ER * Y_d * 10^6 \quad \text{... 7}$$

where OC1 is the fraction of organic carbon content in surface layer of the soil profile (g-C/g-soil), ER is the enrichment ratio (g-C/g-sediment), and Yd is the sediment yield (kg).

Dissolved Oxygen (DO)

Assuming initially that the water from rainfall is saturated with oxygen (100% saturation level), the dissolved oxygen loading from a sub-basin is calculated by subtracting the oxygen uptake by the oxygen demanding substance in the runoff. The amount of oxygen withdrawn from the water depends on the average time of overland flow. The DO loading from a sub-basin is estimated by

$$O_d = O_s - (K_l L_u + \alpha_n \beta_n N) * t_{ov} \quad \text{... 8}$$

where Od is dissolved oxygen concentration (mg/l), Os is the oxygen saturation concentration at temperature Tw (mg/l), K*l* is CBOD deoxygenation rate (day-1), 'alpha_n' is the oxygen uptake rate of organic nitrogen (mg-O/mg-N), 'beta_n' is the oxidation rate coefficient of organic nitrogen (day-1), N is organic nitrogen concentration (mg/l) and tov is the overland travel time (day). For simplicity Kl and 'alpha_n' are assumed to be 1.047 and 'beta_n' is 4.6 (Brown and Barnwell, 1987). The dissolved oxygen saturation concentration is given by

$$O_s = EXP\left[-139.34410 + \left(\frac{1.575701\times 10^{5}}{T}\right) - \left(\frac{6.642308\times 10^{7}}{T^{2}}\right) + \left(\frac{1.2438\times 10^{10}}{T^{3}}\right) - \left(\frac{8.621949\times 10^{11}}{T^{4}}\right)\right] \quad \ldots 9$$

where T is the water temperature in K (K = C + 273.15), and this equation is valid only if T is between 273.15 and 313.15 K (Brown and Barnwell, 1987).

DESCRIPTION OF THE STUDY AREA

We applied the SWAT model with the instream water quality component to simulate the hydrology and water quality in Wister Lake watershed (Figure 1). A detailed description of the watershed and the input data sets for the watershed was presented by Storm *et al*. (1994). The watershed covers approximately 260,000 ha (640,000 acres) streatching between Oklahoma and Arkansas, is situated in the Arkansas river basin. The outlet of the watershed in Wister Lake, situated on the Oklahoma side. Results from a EPA lake survey conducted in 1974 have declared Wister Lake as eutrophic and excessively turbid. A preliminary Total Maximum Daily Load (TMDL) analysis for the Wister Lake watershed was conducted by Smolen et al. (1993), which analyzes the sources of point and non-point source pollution.

Figure 1: Wister Lake Watershed and Subwatersheds

The four primary stream systems flowing into Wister Lake are Poteau river, Black Fork, Fourche Maline, and Holson Creek. There are four continuous stream gage monitoring stations, two on Poteau river, one on Fourche Maline and one on Black Fork. In addition there are three miscellaneous gages having stream flow measurements at approximately six weeks interval. Stream flow data is available from all these stations from December 1991. Along with stream flow data, water quality data is also available from all the seven stations at approximately six weeks interval. Data from all these stations have not yet been consolidated and updated. Therefore, for this work we used the data from only two stations (locations shown in Figure 1). Both of them have continuos stream flow data and six-week water quality data.

The entire watershed was divided into subwatersheds in such a way that the outlet of each subwatershed coincides with one of the stream monitoring stations. Rainfall data is available from five rain gage stations situated in the vicinity of the watershed. Given the locations of the rain-gages, SWAT/GRASS interface picks up the closest rain-gage to each subwatershed. In the process the interface picked up only two of the available rain-gage stations. For the land use map, USGS land use/land cover map of the area was used, and soils map was obtained by digitizing the county soil maps of Leflore and Latimer counties in Oklahoma, and

Scott and Polk Counties in Arkansas. The soil properties were derived from the STATSGO database (USDA-NRCS, 1994). The soils and land use maps are shown in Figures 2 and 3, respectively. The land use map shows that 75 % of the watershed is covered with forest, 23 % with pasture, and the rest of the area is urban, agricultural and rangeland. For this work we used a 100 m resolution for all the data layers though some of the data layers are actually at a lower resolution.

Figure 2: Soils Map of Wister Lake Watershed

Figure 3: Land Use Map of Wister Lake Watershed

SWAT simulations could be conducted based on two subwatershed configurations namely, dominant and virtual basin approach. The details of both approaches are given by Mamillapalli et al. (1996). Due to the low spatial variation in land use and soils we used the dominant approach for this work.

RESULTS AND DISCUSSION

Using the input data accumulated by the SWAT/GRASS interface, the SWAT model was used to simulate the hydrology and water quality in the Wister lake watershed from 1991 to 1994. All the model runs were made starting from 1989 using synthetic weather data for 1989 and 1990, in order to `prime' the model. A minimal calibration of the model for stream flow was conducted. The initial uncalibrated run of the model showed that the model was under estimating stream flow. Therefore, the curve number for all the watersheds was increased by 10 % of the original value and the available water holding capacity of all the soils were reduced by a value of 0.05 mm/mm.

Stream flow

Figure 4a shows the time series plot of cumulative stream flow from 1991 to 1994 for the station at Fourche Maline River at Red Oak (Station 1). In general SWAT over estimated the stream flow, but the simulated trend matches well with the observed. Except for stream flow during November, 1994, all the predicted monthly flows match very well with the observed data. The coefficient of determination between observed and predicted monthly stream flow data is 0.64.

Four-year time series plot of cumulative flow at Poteau River at Cauthron (Station 2) is shown in Figure 4b. At this station the prediction is not as good as Station 1. The predicted total flow for each year is reasonably close to the observed values, but some of the individual monthly values do not match very well. This may be due to lack of good distribution of rain gage locations. Both rain gages that are being used in our simulations are in Oklahoma side of the watershed. But, the stream gage and the subwatersheds flowing through the stream gage are in Arkansas side. Storm et al. (1994) noted significant variation in precipitation within the existing four rain gages in Oklahoma side. Therefore, we feel that for a better prediction of stream flow from Arkansas side, we need weather data from locations closer to the subwatersheds than the existing ones.

Figure 4a: Observed and Simulated Cumulative Stream Flow at Station 1

Figure 4b: Observed and Simulated Cumulative Stream Flow at Station 2

Water Temperature

Figures 5a and 5b show the daily time series plot of simulated water temperature (C) and six-week observed water temperature at Stations 1 and 2, respectively. The breaks in the daily simulated water temperatures are due to the days when SWAT did not predict any flow. The simulated water temperature values are reasonably close to observed data and the trends of the simulations match closely with the observed.

Figure 5a: Observed and Simulated Water Temperature at Station 1

Figure 5b: Observed and Simulated Water Temperature at Station 2

Total Nitrogen

Time series plot of daily simulated total nitrogen concentration and the six-week observed concentration from Stations 1 and 2 are shown in Figures 6a and 6b. The simulated total nitrogen concentrations are unreasonably high compared to the observed concentrations. There may be several reasons for this:

· Looking at the SWAT daily output file (data not presented in this paper) we find that nitrate loading to the stream is high. Since crop production activities in this area is negligible, the main source of nitrogen is biomass residue. The over estimation of total nitrogen concentration in stream water may be due to the over estimation of nitrogen mineralization in the soil.

· SWAT considers default values for observed values of initial soil nitrate and organic nitrogen content in the absence of extensive field observed values. This default value may not be accurate enough.

Researchers agree that modeling nitrogen is one of the most challenging tasks even at field scale. We are trying to find out some more reasons for inaccurate predictions of stream nitrogen concentrations and will try to improve the predictions of SWAT.

Figure 6a: Observed and Simulated Total Nitrogen in Stream Flow at Station 1

Figure 6b: Observed and Simulated Total Nitrogen in Stream Flow at Station 2

Total Phosphorous

Figures 7a and 7b show the simulated daily total phosphorous concentrations and the six-week observed concentrations. We find that SWAT consistently predicts lower total P concentration than the observed values, but the predictions are not very far off from the observed concentrations. Most of the phosphorous in runoff is carried by sediments. Therefore, prediction of sediment in runoff has to be good for reasonable prediction of phosphorus in runoff. In this paper we did not look into the sediment prediction capability of SWAT due to lack of sufficient observed data.

Figure 7a: Observed and Simulated Total Phosphorous in Stream Flow at Station 1

Figure 7b: Observed and Simulated Total Phosphorous in Stream Flow at Station 2

Dissolved Oxygen

Figure 8a shows the daily dissolved oxygen concentration and the six-week observed concentration for Station 1. The breaks in the predicted concentration data correspond to the no-flow days. SWAT predicted lower oxygen concentrations during the high flow periods (October to February) and higher concentration during low-flow periods (April to August). The inaccuracy in reduction of oxygen concentration may be an effect of inaccurate prediction of nutrients. Figure 8b shows the comparison of simulated and predicted oxygen concentration at Station 2. Here the predictions during high flows are reasonably close to the observed, but during low flow periods SWAT tends to over estimate dissolve oxygen concentration in the stream. In general, the dissolved oxygen prediction of SWAT is reasonably good.

Figure 8a: Observed and Simulated Dissolved Oxygen Concentration at Station 1

Figure 8b: Observed and Simulated Dissolved Oxygen Concentration at Station 2

SUMMARY AND CONCLUSIONS

The instream kinetics of QUAL2E was incorporated in a basin scale hydrologic/water quality model, SWAT. We have described the instream water quality component of SWAT and presented the preliminary results from the application of the model to Wister Lake watershed situated in eastern Oklahoma and western Arkansas. The input data for the model were aggregated using a GIS interface. The GIS used by the SWAT interface is GRASS.

SWAT predicted monthly stream at Station 1, situated in Oklahoma side of the watershed with reasonable accuracy. But, the stream flow predictions at Station 2, Which is in Arkansas side is not as good. This may be due to lack of a good distribution of weather stations. The water temperature and dissolved oxygen predictions by SWAT are satisfactory. The nutrient predictions are inaccurate due to many speculative reasons. Efforts for improving the water quality prediction capability of SWAT are under way.

There are only a very few models that are linked with a GIS and capable of simulating both hydrology and water quality on a river basin scale. By adding the instream water quality component for SWAT we have added a new dimension to basin scale modeling. The results from this study are from a `first-cut' preliminary analysis. The GIS-linked SWAT model shows a good potential for being used by a tool to predict the effects of land use activities on surface water bodies.

The SWAT model with instream water quality component can be used as a tool for EPA-recommended TMDL analysis. Having estimates of point source pollution loading from subbasins, SWAT model could be used to analyze the effects of both point and non-point source pollution on the stream body. SWAT can also to used to analyze the effects of alternative management strategies and aid regulatory agencies in decision making.

In summary the major strengths of the model are: (i) simple, yet complex enough to simulate the interactions between weather, crop growth, and land use management on a river basin scale for long periods, (ii) a GIS interface, which will save substantial amount of resources in aggregating input data for large-scale simulations, and (iii) a graphical output interface and analysis tool to visualize the simulation results.

ACKNOWLEDGEMENT

We sincerely thank Dr. Daniel Storm and R. Lakshminarayanan, Biosystems and Agricultural Engineering Department, Oklahoma State University for providing us the GIS map layers.

REFERENCES

Arnold, J.G., P.M. Allen, and G. Bernhardt, 1993. A comprehensive surface-groundwater flow model. J. Hydrol. 142:47-69.

Brown, L. C. and T. O. Barnwell, Jr., 1987. The enhanced water quality models QUAL2E and QUAL2E-UNCAS documentation and user manual. EPA document # EPA/600/3-87/007, Cooperative Agreement # 811883, Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA.

Cluis, D., P. Couture, R. Bégin, and S.A. Visser, 1988. Potential eutrophication assessment in rivers; relationship between produced and exported loads. Schweiz. Z. Hydrol. 50(2):166-181.

Hession, W. C., and V. O. Shanholtz, 1998. A geogrphic information system for targeting non-point source agricultural pollution. J. Soil Water Conserv. 43(3):264-266.

Mamillapalli, S., R. Srinivasan, J. G. Arnold, and B. A. Engel, 1996. Effect of saptial variability on basin scale modeling. To be presented at the Third International NCGIA Conference/Workshop on Integrating GIS and Environmental Modeling, Sante Fe, New Mexico, January, 21-25.

Pelletier, R. E., 1985. Evaluating nonpoint source pollution using remotely sensed data in soil erosion models. J. Soil Water Conserv. 40(4):332-335.

Rewerts, C. C., and B. A. Engel, 1991. ANSWERS on GRASS: Integrating a watershe simulation with a GIS. ASAE papre # 91-2621, ASAE, St. Joseph, MI.

Rosenthal, W. D., R. Srinivasan, and J. G. Arnold, 1995. Alternative river management using a linked GIS-hydrology model. Trans. ASAE, 38(3):783-790.

Smolen, M. D. , R. Lakshminarayanan, W. C. Hession, and D. E. Storm, 1993. Estimating total maximum daily load (TMDL) for wister lake. ASAE paper # 93-2072, ASAE, St. Joseph, MI.

Srinivasan, R., and J.G. Arnold, 1994. Integration of a basin-scale water quality model with GIS. Water Resources Bulletin 30(3):453-462.

Stefan, H. G. and E. B. Preud'homme, 1993. Stream temperature estimation from air temperature. Water Res. Bull. 29(1):27-45.

Storm, D. E., M. D. Smolen, R. Lakshminarayanan, W. C. Hession, 1994. Wister lake watershed project: Annual Report FY 93. Agricultural Experiment Stationand Cooperative Extension Service, Oklahoma State University, Stillwater, OK.

Thomann, R.V. and J.A. Mueller, 1987. Principles of surface water quality modeling and control pp 261-384. Harper & Row Publishers, New York.

Tim, U. S., S. Mostaghimi, V. O. Shanholtz, and N. Zhang, 1991. Identification fo critical nonpoint pollution source area using geographic information systems and simulation modeling. ASAE paper # 91-2114, ASAE, St. Joseph, MI.

USDA-NRCS, 1994. State Soil Geographic Data Base, US Department of Agriculture - Natural Resource Conservation Service, National Soil Survey Center, Publication # 1492, Washington, DC.

Tharacad S. Ramanarayanan
Research Associate
Blackland Research Center
Phone: (817) 770-6680
Email: rama@brcsun0.tamu.edu

Raghavan Srinivasan
Associate Research Scientist
Blackland Research Center
Phone: (817) 770-6670
Email: srin@brcsun0.tamu.edu

Jeffrey G. Arnold
Hydraulic Engineer
USDA-ARS
Phone: (817) 770-6502
Email: arnold@brcsun0.tamu.edu

Address: 808 East Blackland Road, Temple, TX 76502
Fax: (817)-770-6561

Jennifer L. Dungan and Joseph C. Coughlan

Redefining the spatial support of environmental data in the Regional HydroEcological Simulation System

---

ABSTRACT

The acquisition of spatial data in a GIS usually brings together data defined on variable and/or inconsistent support. These data are problematic for models that operate on them, especially if these models were built from data relationships defined at the same support. This is the "modifiable areal unit problem" (Openshaw 1984) of environmental modeling and GIS and is a component of uncertainty in any model result. Common practice is to ignore support differences and use data sources naively.

We provide an example of the problem and a proposed solution using the Regional HydroEcological Simulation System (Coughlan and Dungan in press). The problem is manifested in the model input data: temperature, radiation and precipitation defined at points; leaf area index defined by remotely sensed images and soil water holding capacity defined by irregularly-shaped polygons interpreted from soil survey. As a result, processes represented by model equations will be applied to combinations of variables that represent different spatial locations and areas, reducing the advantages of utilizing a mechanistic approach.

The ideas of probability field simulation can be applied to this problem in an attempt to represent information on a common support and simultaneously characterize the uncertainty inherent in using spatial models that are different for every model variable.

INTRODUCTION

All measurements have a support, that is an area or volume for which they pertain. For example, a soil core has a support equal to the volume of the core; an infrared gas analyzer for measurement of photosynthesis has a support equal to the area or volume of a single leaf. Simulation model predictions also have a support, since they are based on measurements and attempt to quantify rates or states that are not easily measurable. While the support of measurements is often not specifically addressed or ill-defined, the support of a simulation model is even less often recognized. This despite the fact that a cornerstone of physically-based process modelling is that values of model variables and parameters represent some area or volume in space.

In the ideal case, all measurements would be made at the same location and be based on the same support, the model would accurately capture the important processes occurring within that area or volume, and therefore the model prediction would apply to that support. The ideal case is rarely met, and this results in a source of error in model results that is difficult to quantify. In some ways, this problem is related to the modifiable areal unit problem (M.A.U.P.) central to geographic research (Openshaw 1984, Fotheringham and Wong 1991). The problem states that changing the shape (the "zoning problem") or size (the "aggregation problem") of the units on which data are mapped can change the resulting correlations or statistical models generated from the data.

Because of the different sources and measurement technologies of environmental data used in simulation models, the supports of these data can vary widely. The data models and interpolation procedures used to create spatial fields from these data also affect the support. The raster data model, though usually based on a fixed size for each spatial unit represented, does not guarantee a consistent support for the data it describes. The vector model involves varying size and shape of the spatial units and therefore can never represent a fixed support. The resulting condition is that the covariance among input data sets can be greatly affected by choices of data models and interpolation algorithms rather than by the data themselves.

Is this modifiable areal unit problem a large source of error in ecosystem model predictions? To prepare to investigate this question, we propose a method for redefining the spatial support of one database, a soils layer, to make it compatible with the support of a grid-based vegetation layer. These two data layers are critical for many deterministic ecosystem models, including the Regional HydroEcological Simulation System (RHESSys, see Nemani et al 1993, Coughlan and Dungan (in press)), which we will use through the paper as a reference example.

APPROACH

Like many environmental models operating in the spatial domain, RHESSys is based on one-dimensional, deterministic equations that are parameterized over space to yield two-dimensional results. The equations themselves do not have a spatial component, but since the values of equation input variables vary spatially, the output variables do as well. Each input variable can be potentially described as a spatial data layer; in practice only some variables are described this way and the rest are assumed to be constant over space. To make predictions with RHESSys, all spatial input variables should be described on a consistent support -- the distances, areas or volumes that the model equations pertain to.

Requirements of the RHESSys model

Spatial input variables that are used in RHESSys include vegetation, terrain, climate and soil quantities (Figure 1). The green surface active in evapotranspiration and photosynthesis is described using leaf area index (LAI), which ranges from 0 to approximately 12 in units of m^2 / m^2 . LAI controls interception of precipitation, is used in a Penman-Montieth formulation describing transpiration, and influences photosynthetic rate in several ways including by radiation interception. Incoming solar radiation is also mediated by terrain, described by a submodel that uses slope and aspect. Air temperature is important in influencing the rates of transpiration, snowmelt and photosynthesis. The soil substrate is modelled as a water reservoir with a specific capacity, tapped by water demands of the canopy, receiving inputs from precipitation and creating runoff in cases of overflow.

Figure 1. *Data and model structure of RHESSys*

Data Sources

RHESSys and related models have been parameterized at a variety of scales (Band et al. 1993, Running et al. 1989, Running and Hunt 1993, Pierce and Running 1995). In general, vegetation has been derived from optical remote sensing data; climate has been interpolated from meteorological station data, terrain is described by digital elevation models, and soils data from vector-based soil maps. We describe the set of sources that has been used for regional-scale simulations:

1. LAI is estimated from visible and near infrared data from the Advanced Very High Resolution Radiometer (AVHRR) sensor. The algorithm of Running et al. (1989)

   is used to predict LAI at every 1.1 x 1.1 km grid cell based on the normalized difference vegetation index (NDVI). Though grid cell size is an approximation to the point spread function of the sensor, this represents an approximately consistent support for model simulations. This algorithm produces an LAI "image" (Figure 2).

   Figure 2. *Leaf Area Index estimated using AVHRR NDVI*

2. Elevation, slope and aspect are extracted from the 1 arc second terrain model provided by the U.S. Defense Mapping Agency. The production of this data layer and its subsequent aggregation to a 1.1 km grid involves some smoothing (a reduction in extreme values). Bindlish and Barros (in press) and have effectively addressed this problem with a fractal interpolation.
3. Precipitation, solar radiation and maximum and minimum daily temperatures are based on meteorological stations at a quasi-point support. Data from the DEMs are used to estimate values between measurements, which should result in a support similar to that of elevation.
4. Soil available water capacity (AWC) is a quantity provided in the State Soil Geographic Data Base, or STATSGO (Soil Conservation Service, 1991), which was designed for regional applications (Reybold and TeSelle, 1989). In STATSGO, soil mapping units, irregularly shaped polygons described by a vector model, comprise the nominal support of AWC, despite the fact that AWC measurements may have been made at a much smaller support. Users of the maps must assume that each value is representative of the entire soil unit, because sub-unit information is no longer available. Each soil map unit in STATGSGO actually includes ranges of values, and proportions of component map units that are below the STATSGO scale. Figure 3 shows the soil units from the Oregon

STATSGO database.

Figure 3. *Soil map units in the Oregon STATSGO database. Polygons colors are randomly assigned and do not reflect values of STATSGO attributes.*

Support definition

Given that all data sources are available at different, sometimes variable supports, to what area does a simulation pertain? Common practice is to ignore the support differences, and aggregate or disaggregate each source naively. The practice includes rasterizing polygon data, averaging or resampling to a specific grid. The 1.1 x 1.1 km^2 support was chosen as the grid for RHESSys simulations. LAI is defined on this support because it is derived from the AVHRR data on this grid. Terrain data area aggregated from 100 x 100 m^2 to 1.1 x 1.1km^2 using simple averaging. The supports of soil and climate variables are less consistent. For the purposes of this paper, problems of climate interpolation are ignored and we concentrate on the soil information. The soil mapping units range in size from .02 to 12,000 km^2, comprising a minimum of .01 to a maximum of 10,500 grid cells. Since soil water capacity is not available at this support, we can represent our uncertainty about the soil value that exists at each grid cell at the chosen support by a distribution (which can be thought of as a histogram, a probability density function, or its integral, the cumulative distribution function, CDF).

Constructing CDFS of AWC

In the limit of no soil information, the cdf at each grid cell would be a straight line, defining a uniform distribution -- every physically possible soil water holding capacity would be equally likely and there would be no geographic variation in the cdfs (Figure 4a). At the other extreme, we would have error-free knowledge of the average available water capacity for each 1.1x1.1 km grid cell (Figure 4b). Our knowledge about the soil is actually intermediate between these two extremes, and that knowledge comes from at least three sources:

Figure 4. *a) Cumulative distribution function for available water capacity in the absence of information, b) Cdf for AWC with perfect knowledge.*

1. The STATSGO database. Each polygon in the database may contain a number of components, equivalent to soil series or phases, each of which is characterized by a set of soil parameters. Some of these parameters are given as ranges instead of single values. For example, one layer of a component may be characterized by a minimum available water capacity of .05 inches/inch and a maximum AWC of .9 inches/inch. The locations of the components are not specified, but their proportions are.
2. The vegetation amount. The leaf area index supported by the soil at a cell tells us something about the water relations at that cell. If there is a low LAI, that tells us little about the soil (does not narrow the cdf). If there is a high LAI, that tells us that soil water holding capacities are likely to be at the upper end of their range.

3. The terrain. The cdfs may be further constrained with elevation information, given some model of soil development (Zhu and Band 1995).

These three sources of information can be used to construct cdfs for each grid cell used in a RHESSys simulation. We considered the first two in this paper.

Construction of an AWC cumulative distribution function for each grid cell defined with the chosen support can be framed as a probability problem, where we are trying to update the *prior* probability about the available soil water capacity given by the soil map with information about leaf area index to constrain that prior probability. This posterior probability is a conditional probability; the probability that the available soil water capacity is less than or equal to a certain value given the LAI is greater than or equal to a certain value. Using the notation

The problem can be solved using Bayes' rule:

We can define P(A) using the information from the soil map (see prior CDFs). P(B) can be defined using the remote sensing information, and P(B|A) can be defined using ecological reasoning (see conditional CDFs).

Constructing prior CDFS from STATSGO

Each map unit polygon in the STATSGO database may contain up to 21 components. The locations of the components are not specified, but their proportions are. In addition, each component is described by a range of soil water capacities rather than a single value. Minimum available water capacity **awcl** and maximum available water capacity **awch** are given for each soil layer or horizon expressed as inches/inch. Since the RHESSys soil model does not include layers, the **awcl** and **awch** values are integrated over top soil layers to 1 m. (One m is the assumed rooting depth used in RHESSys simulations):

where *k* is the index of the soil layer, and *N_k* is the number of soil layers. If the last layer extends beyond 1 m, the AWC value for that layer is weighted by the proportion of that layer that extends to 1 m.

The combination of component data into a single cdf can be illustrated with an example. Say there are five components in a mapping unit. They are described as follows:

1. AWC values range from .05-.07 and the component makes up 35% of the unit.
2. AWC values range from .08-.15 and the component makes up 15% of the unit.
3. AWC values range from .15-.5 and the component makes up 20% of the unit.
4. AWC values range from .2-.3 and the component makes up 20% of the unit.

5. AWC values range from .35-.4 and the component makes up 10% of the unit.

Figure 5a shows the cdfs for each component.

Figure 5. *a) Cdfs from each of the components making up the soil unit. The variable Z represents AWC. b) Cdf of AWC for the whole map unit.*

To define the global CDF for the mapping unit the component CDFs are added. (Since they are independent events, their union is equal to the sum of their probabilities.) This is the prior cdf, P(A), shown in Figure 5b.

Constructing conditional CDFs using ecological knowledge

The following ecological reasoning can be applied to the problem. At any location, available soil water capacity must be adequate to sustain the LAI found there. This logic can be used to model the probability that LAI is greater than or equal to a given value given that AWC is less than or equal to a given value P(B|A). If soil water is less than .1, there is a low probability that LAI is high. Nemani and Running's (1989) description of description of hydrologic equilibrium conditions for mesic sites were used to define AWC values required by a given LAI. This logic leads to a suite of P(B|A) conditional probabilities, one for each LAI value. Three are shown in Figure 6. A smooth surface was fit through the models to account for all LAIs.

P(B) is assumed to be 1, though uncertainty about the remote sensing algorithm could be included. The posterior cdf is then obtained by multiplying P(A) with P(B|A) at each grid cell.

Figure 6. *Conditional cdfs defined at three LAI values.*

Drawing from cdfs

Once the cdfs are defined for each grid cell, the remaining task is to draw from each one to generate a soil value for each cell. There are many ways this could possibly be done, but one way that has been used in other earth science applications is called probability field simulation (Srivastava 1992). A probability field simulation is simply an image, defined on our target grid, with a value between 0 and 1 allowing us to draw from the cdfs that we have defined. The reason this is different from a random draw is because of *geography* -- i.e. that values close together will be more similar that those farther apart. That means that if we have drawn a value at one cell, the neighboring cells are more likely to have a value similar to it. The perfectly random probability field (the image would look like white noise) would describe the most conservative description of our uncertainty. While it is difficult to specify an autocorrelation function, it is reasonably certain that the perfectly random probability field is not a good model.

For selecting from AWC cdfs, probability fields were generated using unconditional simulation specified with a spherical variogram model with a range of 22 km, an arbitrary model used here for the purposes of demonstration. The p value of each cell is used in the quantile algorithm to obtain a value of AWC at that cell. Three probability fields are shown in Figure 7.

Figure 7. *Three probability fields.*

RESULTS

Three realizations, or possible soil maps, defined on the support grid and obtained using the probability fields in Figure 7 are shown in Figure 8. The spatial patterns below are very different to that of a STATSGO soils map of Oregon (compare with figure 1). Each of the three maps are completely consistent with the information in the STATSGO database. But unlike the STATSGO soil unit map they are at the same support as the LAI information. The spatial patterns in the realizations appear to be strongly governed by the probability fields, indicating the importance of choosing a good spatial autocorrelation model.

Figure 8. *Three realizations of the soil map showing available water capacity in inches/inch*

CONCLUSIONS

This paper has described a proposed method to represent soils and vegetation information on a common support and simultaneously characterize the uncertainty inherent in a soils database. The method is similar to bootstrapping (Foussereau et al 1993, Journel 1995) used in other earth sciences. It includes an additional constraint of LAI which serves to increase the covariance between AWC and LAI in the resulting input layers, thereby narrowing the range of uncertainty over univariate bootstrapping methods. One important research question is the rational definition of spatial autocorrelation models for the probability fields. The measurement of the spatial dependence of soils variables, particularly over large regions, would address this question. Such information in not currently availabe in regional soils maps.

We plan to use these alternative soil maps in RHESSys and compare the resulting output fields with those based on average values from soil units (Lathrop et al. 1995). By doing so, we hope to address the uncertainty added by the M.A.U.P. in the simulation of photosynthesis and evapotranspiration regionally.

ACKNOWLEDGEMENTS

This research has been supported by NASA's Mission to Planet Earth.

REFERENCES

Band, L.E., P. Patterson, R. Nemani and S.W. Running (1993) Forest ecosystem processes at the watershed scale - incorporating hillslope hydrology. Agricultural and Forest Meteorology 63: 93-126.

Bindlish, R. and A.P. Barros (in press) Aggregation of digital terrain data using a modified fractal interpolation scheme. Computers & Geosciences.

Coughlan, J. and J. Dungan (in press). Combining remote sensing and forest ecosystem modeling: An example using the Regional HydroEcological Simulation System (RHESSys) IN The Use of Remote Sensing in the Modeling of Forest Productivity at Scales from the Stand to the Globe, ed. H.L. Gholz, K. Nakane and H. Shinoda, Kluwer Academic Publishers, Dordrecht.

Fotheringham, A.S. and D.W.S. Wong (1991) The modifiable areal unit problem in multivariate statistical analysis Environment and Planning A 23:1025-1044.

Lathrop R.G., J.D. Aber and J.A. Bognar (1995) Spatial variability of digital soil maps and its impact on regional ecosystem modeling Ecological Modelling 82:1-10

Nemani, R.R. and S.W. Running (1989) Testing a theoretical climate-soil-leaf area hydrological equilibrium of forests using satellite data and ecosystem simulation, Agricultural and Forest Meteorology, 44:245-260.

Nemani, R.R., S.W. Running, L.E. Band and D.L. Peterson (1993) Regional hydro-ecological simulation system: An illustration of the integration of ecosystem modeling with a GIS. pages 296- 304. in:Environmental Modeling with GIS. M.F. Goodchild, B.O. Parks and L.T. Steyaert eds. Oxford University Press.

Openshaw, S. (1984) The Modifiable Areal Unit Problem IN Concepts and Techniques in Modern Geography Number 38. GeoBooks, Norwich.

Pierce, L.L. and S.W. Running (1995) The effects of aggregating sub-grid land surface variation on large-scale estimates of net primary production. Landscape Ecology, 10:239-253.

Reybold, W.U and G.W. TeSelle (1989) Soil geographic data bases, Journal of Soil and Water Conservation, 44:28-29.

Running,S.W. and E.R. Hunt Jr. (1993) Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. pages 141-158 in: Scaling Processes Between Leaf and Landscape Levels. J.R. Ehleringer and C. Field eds. Academic Press.

Running, S.W., R.R. Nemani, D.L. Peterson, L.E. Band, D.F. Potts, L.L. Pierce, and M.A. Spanner (1989) Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystem simulation. Ecology 70:1090-1101.

Soil Conservation Service (1991) State Soil Geographic Data Base (STATSGO): Data Users Guide. SCS Miscellaneous Publication Number 1492, Washington, D.C. 88 pages.

Srivastava, R.M. (1992) Reservoir characterization with probability field simulation, in Annual Technical Conference of the Society of Petroleum Engineers, SPE 24753, Washington, D.C., pp. 927-938.

A. Zhu and Band, L. E. (1994) A knowledge based approach to data integration for soil mapping. Canadian Journal of Remote Sensing, 20:408-418.

Jennifer L. Dungan
JCWS, Inc.
MS 242-4
NASA Ames Research Center
Moffett Field, CA 94035-1000
***Email**: jdungan@gaia.arc.nasa.gov*

Stephan J. Maas and Paul C. Doraiswamy

Integration of Satellite Data and Model Simulations in a GIS for Monitoring Regional Evaporation and Biomass Production

A method is described for estimating regional evaporation and plant primary production. The method employs a relatively simple model to simulate evaporation and biomass production based on ambient environmental conditions. The model containd submodels for the vegetation growth and soil water balance processes. Data obtained from NOAA AVHRR imagery are used to calibrate model simulations. Spatially distributed data and model simulations are organized within a raster GIS. The procedure is demonstrated using data from the Walnut Gulch Experimental Warershed, a semiarid rangeland located in southeastern Arizona.

INTRODUCTION

Satellite remote sensing can be an effective method for assessing surface evaporation and vegetation resources of a geographical region (Jackson 1985, Moran and Jackson 1991, Moran et al. 1992, Kustas et al. 1994). While satellites can easily survey large areas of the earth's surface, a drawback to their use in operational monitoring programs is the relative infrequency of their observations as a result of their overpass frequency or the occurrence of cloud cover. Satellite observations represent discrete time events which may indicate little about how the biosystem got to its observed state or what its condition will be in the future. In contrast, mathematical models of the vegetation-soil-atmosphere system can produce essentially continuous descriptions of vegetation growth and evapotranspiration (Maas et al. 1992, 1993). They can also be used to project the future condition of the biosystem. Unfortunately, deterministic simulation models utilize a set of parameters and inputs that is specific to the location being simulated. The resulting model simulation is appropriate only for an area within which the values of parameters and inputs do not significantly change. This precludes the explicit application of simulation models to a geographical region in which there is significant spatial variation in environmental conditions.

Previous studies (Maas 1988a, 1988b, 1991a, 1991b, Maas et al. 1992, 1993, Moran et al. 1995) have demonstrated that remotely sensed information can be incorporated into model simulations of agricultural and natural biosystems. These efforts have involved the extraction of remotely sensed data for a point in the landscape for which the model was applied. In this presentation, we describe the use of a geographic information system (GIS) to expand the application of a simulation model over a geographical region to make use of the inherent spatial quality of remotely sensed image data. This procedure is demonstrated using data acquired during 1990 at the Walnut Gulch Experimental Watershed, a semiarid rangeland located in southeastern Arizona.

OBJECTIVES

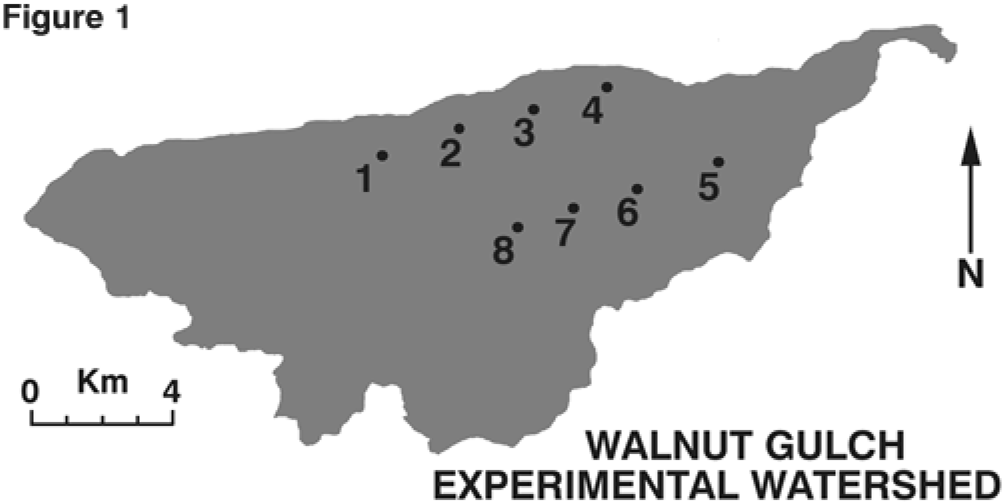
- To demonstrate the incorporation of remotely sensed information in model simulations
- To demonstrate the organization of data and execution of the simulation model in a GIS

APPROACH

Field Study

Field data were collected in 1990 from sites within the Walnut Gulch Experimental Watershed located in southeastern Arizona (see Figure 1). Micrometeorological observations, including air temperature and solar radiation, were collected at eight sites within the watershed. Aboveground biomass, percent canopy cover, and soil moisture were measured periodically at these locations during the study. Data collection procedures are summarized by Kustas et al. (1991a). An essentially uninterrupted sequence of 27 days of data (July 28 through August 15) was available for use in this modeling effort.

Figure 1



The Model

The model used in this study is similar to earlier versions described by Maas et al. (1992, 1993) and Moran et al. (1995). It consists of two submodels-- a soil water balance submodel and a vegetation growth submodel. These submodels produce daily simulations of surface evaporation, soil moisture, leaf canopy density, and aboveground biomass production.

The formulation of the soil water balance submodel is based on the following assumptions:

1. For vegetated surfaces in semiarid environments, the contribution of soil evaporation to

evapotranspiration (ET) is negligibly small compared to the contribution from plant transpiration, except immediately after a rainfall
2. When soil water is abundant, ET approaches potential ET (PET) for the vegetation canopy
3. Except in areas of steep slope, surface runoff of rainfall occurs following saturation of the surface soil layer

Based on these assumptions, ET is determined on most days by the degree to which the evaporation from the vegetation canopy approaches PET and the degree to which the vegetation canopy covers the surface. On days immediately following a rainfall, a contribution from soil evaporation to ET may also occur.

Studies involving agricultural crops indicate that the ratio of vegetation canopy evaporation to PET can be expressed as a function of the available soil water in the plant root zone. Available soil water is defined as the amount of water between the wilting point of the vegetation and the maximum drained capacity of the soil, normalized by the maximum drained capacity. Thus, available soil water (ASW) is a number between zero and one. Studies also indicate that, when soil water is abundant, the ratio of vegetation canopy evaporation to PET increased with increasing leaf area index (LAI). LAI is a measure of vegetation canopy density, determined as the ratio of total plant leaf area to plant ground surface area. When soil water is not limiting, vegetation canopy evaporation approaches PET at an LAI of around 3 (roughly complete vegetation canopy cover).

Using these relationships, the soil water balance submodel calculates vegetation canopy evaporation (TRAN) as follows,
[Eq.1] TRAN = PET*FSW*FGC
where FSW and FGC are the ratios determined by ASW and LAI, respectively. On the day following a rainfall, soil evaporation (SOIL) is calculated as follows,
[Eq.2] SOIL = (PET-TRAN)*(1-FGC)
The value of SOIL cannot exceed the amount of rain that fell on the previous day. The value of ASW on any day during the simulation is given by the balance,
[Eq.3] ASW = ASW' + RAIN - TRAN - SOIL - RUNOFF
in which ASW' is the soil water on the previous day and RAIN is rainfall. The value of RUNOFF is determined as the positive difference between the term ASW'+RAIN-TRAN-SOIL and the maximum drained capacity of the soil. Admittedly, this is a simple model, but it suffices for the demonstration contained in this presentation.

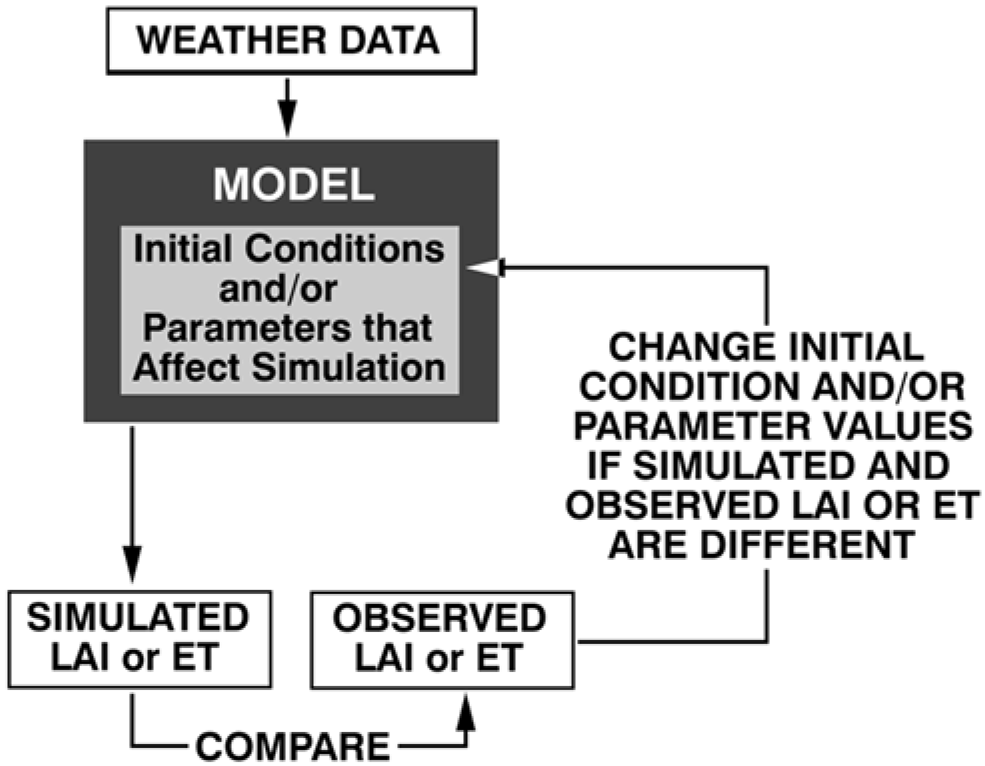
The formulation of the vegetation growth submodel is similar to that used in earlier agricultural crop growth models (Maas 1992, 1993a, Maas et al., 1993, Moran et al. 1995). Daily biomass growth (DELBM) is determined as follows,
[Eq.4] DELBM = APAR*EC*FTEMP*FSW
where APAR is the photosynthetically active solar radiation absorbed by the vegetation canopy, EC the "conversion efficiency" between APAR and new biomass, and FTEMP a nondimensional function of ambient temperature. APAR is determined as a function of LAI. Newly formed aboveground biomass is partitioned between leaves and stems, and existing leaf area senesces at a rate determined by its age. The presence of the factor FSW in Equation 4 results in a decrease in new biomass production with decreasing soil water.

Model Calibration

Remotely-sensed information is not required by the model to simulate surface evaporation and biomass production. Remotely sensed data allow the model simulation to be operationally adjusted during the simulation period to insure that modeled and observed conditions are in agreement. Such adjustment is often neccessary because it is recognized that the models are only mathematical approximations of the real system. Maas (1988a) showed that the most effective method of adjusting simulation models was through reinitialization and/or reparameterization. In these procedures, which are collectively termed "within-season calibration", the values of certain model initial conditions and/or parameters are manipulated until the model simulation of a quantity fits a corresponding set of observations. An iterative numerical procedure is built into the model to manipulate the initial conditions and/or parameters so that they converge on values that result in a fit of the simulation to the observations. The mathematics of this numerical procedure has been described in detail by Maas (1992, 1993b). Maas (1991a, 1991b) showed that within-season calibration using remotely sensed data could significantly improve the accuracy of agricultural crop growth simulation models.

Two quantities in the soil water balance submodel are affected by the calibration procedure-- the maximum drained capacity of the soil, and the initial amount of soil water at the start of the simulation. Calibration of these quantities occurs through comparison of modeled ET to obervations of ET. In the vegetation growth submodel, two other quantities take part in the calibration procedure-- the initial value of LAI at the start of the simulation, and a parameter that determines the lifespan of leaves following their formation in the canopy. For these quantities, calibration occurs by comparing simulated LAI to observations of LAI obtained during the simulation period. This calibration procedure is shown diagrammatically in Figure 2. In the complete model, calibration is first performed for the vegetation growth submodel, using default values for the soil water related quantities. The resulting LAI simulation is then used in a calibration of the soil water balance submodel. The resulting soil water conditions are used in a second calibration of the vegetation growth submodel. This sequence continues until changes in the vegetation and soil water simulations from subsequent calibrations become negligibly small. Usually two to three passes through the calibration sequence are required for this convergence.

Figure 2



WITHIN-SEASON CALIBRATION

Remotely Sensed Information

NOAA-11 AVHRR images were obtained for two dates during the 1990 field study, July 28 (Day 209) and August 4 (Day 216). As the study period occurred during the relatively rainy Arizona "monsoon" season, these were the only days for which the watershed region was cloud-free. Atmospheric correction and processing of the satellite images is described by Kustas et al. (1994). The images in band 1 (red waveband), band 2 (near-infrared waveband), and bands 4 and 5 (thermal infrared wavebands) were resampled to create 1.1-km pixels, and portions of the images containing the watershed were extracted an coregistered. Pixels falling outside of the watershed boundary were masked. A split window approach (Doraiswamy and Perry 1991) was used to determine the surface temperature for each pixel from data in the thermal infrared bands. Data in the red and near-infrared bands were used to compute the normalized difference vegetation index (NDVI) for each pixel according to the formula,
[Eq.5] NDVI = (NIR-RED)/(NIR+RED)
in which NIR and RED are the surface reflectances in the near-infrared and red wavebands.

Data Organization (GIS)

The surface temperature and NDVI data obtained from the NOAA AVHRR imagery is in a spatially distributed form. A GIS was employed to produce spatially distributed simulations for comparison with the remotely sensed information. Thus, the remotely sensed data could then be used in calibrating the model simulations.

For this application, we chose to use a raster GIS to organize the spatial arrays of the environmental conditions (daily air temperature, solar irradiance, PET, and rainfall) needed to produce a model simulation. The scale of these arrays matched that of the remotely sensed data. If one then accepts the assumption that environmental conditions do not vary significantly over the area of a grid cell, executing the model for each grid cell in the array produces the desired spatially distributed simulation. The assumption of homogeneity within the grid cell can almost always be satistified by using relatively small array elements. In practice, the user must decide on an array element size that optimizes the balance between known spatial variability and desired simulation accuracy. Since the Walnut Gulch watershed has a naturally vegetated, gently rolling landscape, the use of 1.1-km array elements should be acceptable for this demonstration.

Observations of weather data required for determining daily air temperature, solar radiation, and PET were obtained from the eight micrometeorologocal stations located within the watershed. Daily PET at each station was calculated using the method described and validated by Van Bavel (1966). Values of daily air temperature, solar radiation, and PET were extrapolated from the eight micrometeorological station sites to the center of each grid cell using a distance-weighted procedure,

[Eq.6] Vx = [SUM (Vi/Di)]/[SUM (1/Di)]

where Vx is the data value computed for the grid element, Vi the observed data value at station i, Di the distance between the center of the grid element and station i, and SUM (...) indicates a summation of the values in the parentheses over the eight stations. This resulted in the generation of a spatial array (essentially, an "image") of daily temperature, solar radiation, and PET for each of the 27 days in the study. Because these three factors are continuous over the lanscape, and the Walnut Gulch watershed is relatively small (27 km long and 11 km wide), the values of each factor showed little spatial variation on any given day.

Rainfall was observed at 112 sites in and around the watershed. Equation 6 was used to compute spatial arrays of daily rainfall, except that only observation sites within 5 km of the center of a grid element were used in the extrapolation. This resulted in the generation of a daily rainfall array (or "image") for each of the 27 days in the study that preserved the discontinuous quality of that factor.

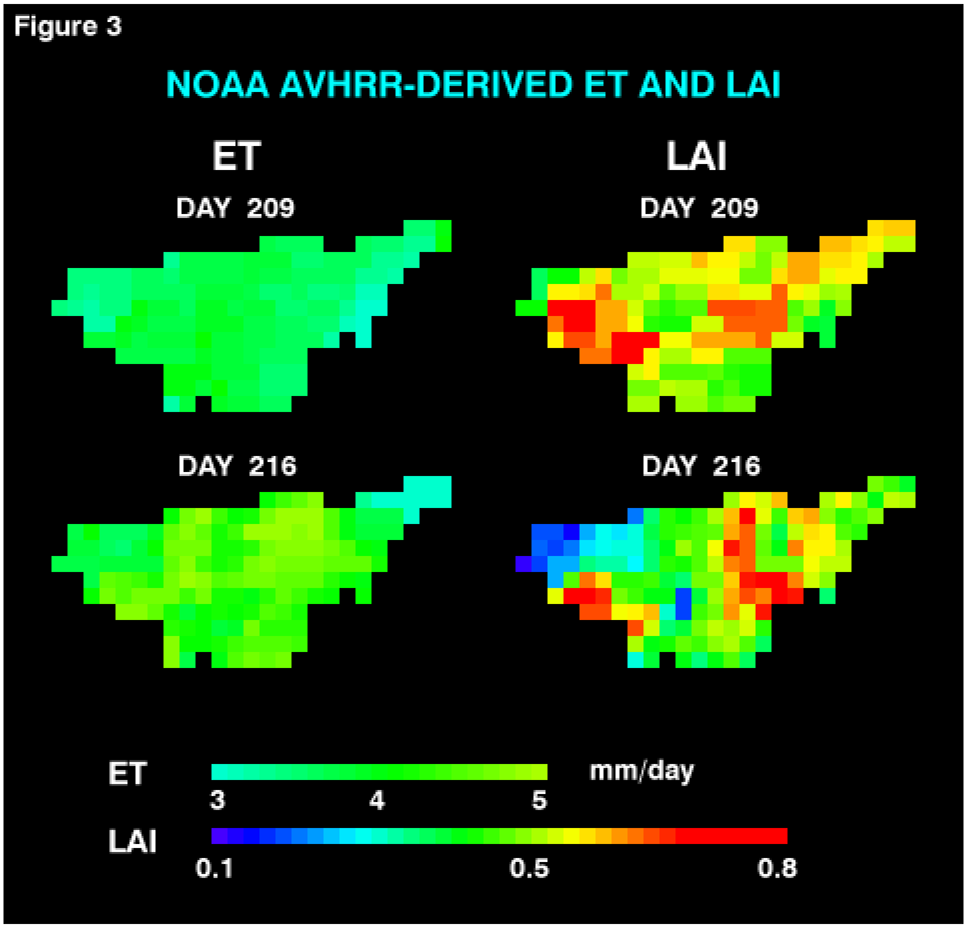
The generation of these arrays provided the environmental data needed to execute the model and produce a simulation of ET and LAI for each grid element. The remotely sensed surface temperature and NDVI images for Day 209 and Day 216 were used to produce spatial arrays of observed ET and LAI for calibrating the model simulations for each array element. ET was estimated from surface temperature using the following relationship (Kustas et al. 1991b),

[Eq.7] ET = RNET - G - [A -B*(TSURF-TAIR)]

in which RNET is net radiation, G the heat flux into the ground, TSURF the remotely sensed surface temperature, and TAIR the air temperature. The constants A and B have the values 1 and 0.2, respectively, and the units of RNET and G are mm/day (evaporated water). RNET and G were evaluated from data observed at the eight micrometeorological stations. Since the values of RNET, G, and TAIR exhibited little spatial variation over the watershed, an average value of each was used in Equation 7 in computing ET for each array element on Day 209 and Day 216. LAI was computed from NDVI for each array element using the following relationship,

[Eq.8] LAI = 4.147*NDVI - 0.276

This relationship was developed from ground-based reflectance and ground cover observations made at sites within the watershed. The resulting ET and LAI "images" for Day 209 and Day 216 are shown in Figure 3.



A number of raster GISs are available for applications like that of this study. A feature of the GIS used in this study (EASI/PACE by PCI, Inc., Ontario, Canada) was its incorporation of a programming language that allowed us to combine data management, image display, and model execution in a single program. The user has a wide variety of image- and data-related procedures within the system that can be incorporated into a program, along with standard I/O, looping, and branching features and the ability to call, execute, and return from external programs (like the model used in this study). The program activities carried out for this demonstration were relatively simple-- for each array element, extract the environmental data needed to run the model, extract the remotely sensed ET and LAI data, execute and calibrate the model, and build spatial arrays of the desired model results (ET, LAI, soil water, and biomass). Following completion of the program, the daily "images" of ET, LAI, soil water, and biomass could be viewed in sequence (like a movie) to visualize the changes in the watershed over time.

RESULTS

The model appeared to produce results consistent with what was observed on the ground. ET measured on 10 days at the Kendall micrometeorological site (station #5) averaged 2.49 mm/day (std. dev. 1.06 mm/day). Modeled ET for the array element containing this site over the same period averaged 2.50 mm/day (std. dev. 0.75 mm/day). The complete ET simulation for this array element is shown in Figure 4. Simulated LAI and biomass growth for this array element are shown in Figure 5. The biomass simulation is in reasonable agreement with the magnitude of biomass observations obtained around the Kendall micrometeorological site (the simulation does not include dead biomass already in the vegetation canopy at the start of the simulation). Daughtry and Perry (1991) reported an LAI of 0.6 from measurements around the Kendall site during the period from Day 207 to Day 217. This is in reasonable agreement with an average simulated LAI of 0.57 (std. dev. 0.04) over that period for the array element containing that site.

Figure 4

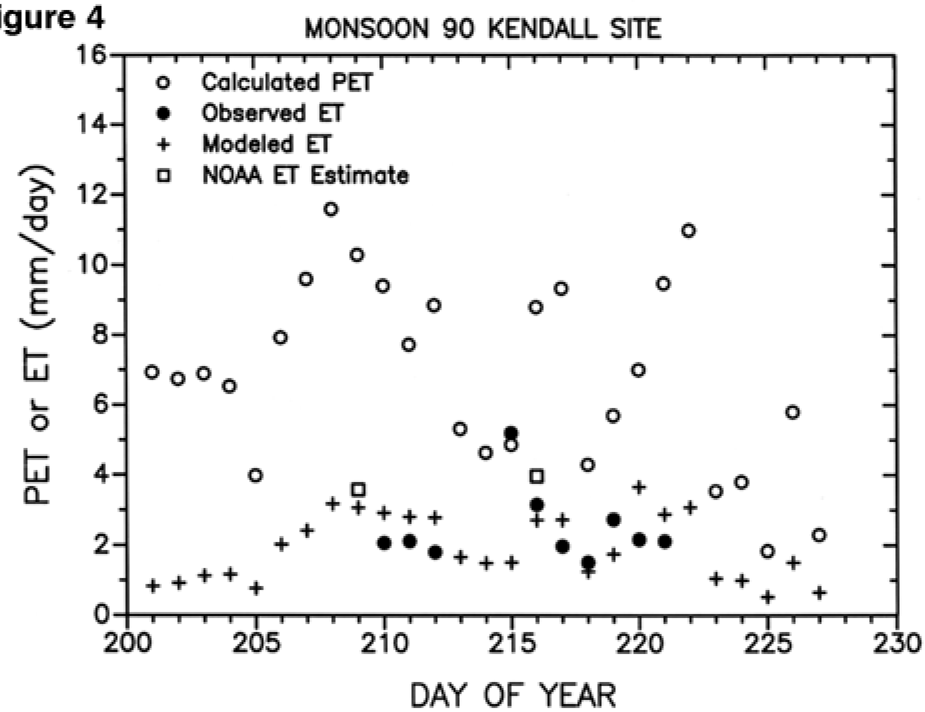
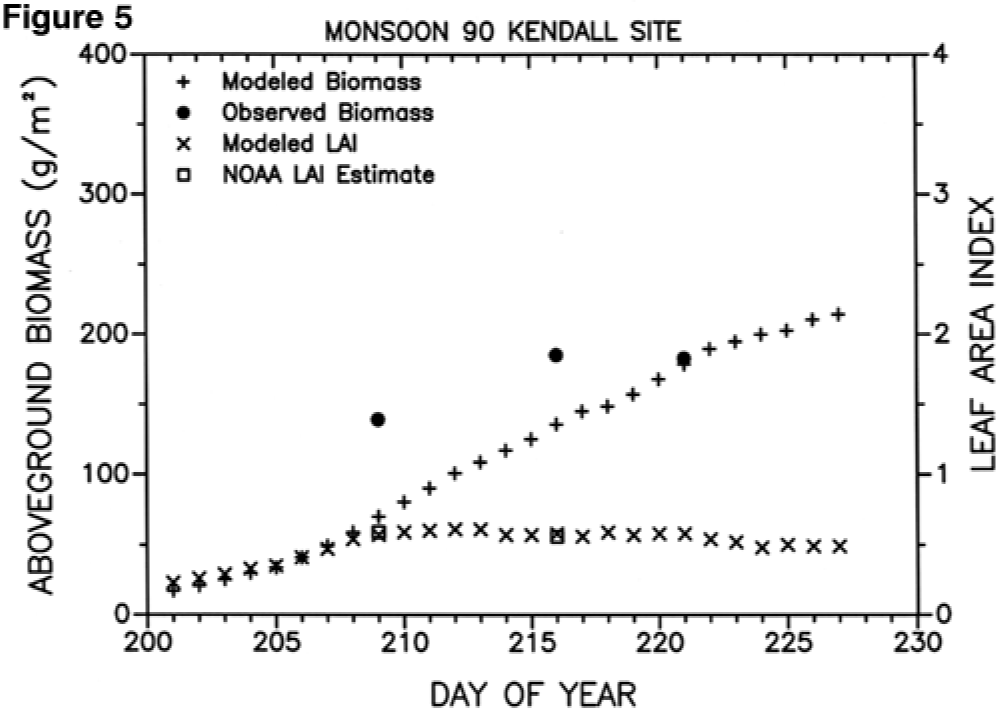
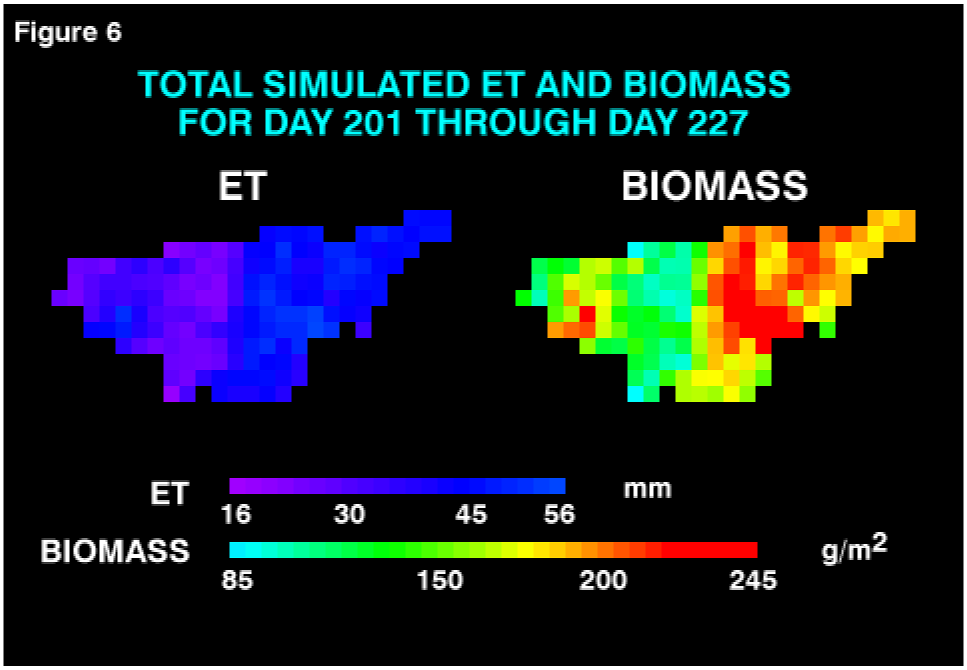


Figure 5



Total ET and biomass production "images" produced by the model for the 27-day period of this study are shown in Figure 6. Total water loss through evaporation from the entire watershed is estimated at 7.56 gigaliters (613,000 acre-ft) over the duration of the study. This may be compared with the amount of water input to the watershed through rainfall, estimated to be 18.32 gigaliters (1,485,000 acre-ft). It is estimated that a total of 32.75 gigagrams (29,700 U.S. tons) of new aboveground biomass was produced within the watershed during the study period. Assuming a carbon content of 38 percent of the biomass carbohydrates (Charles-Edwards 1982), this would imply that 12.45 gigagrams (11,290 U.S. tons) of atmospheric carbon would have been fixed by vegetation within the watershed during the 27 days of this study.



CONCLUSIONS

This relatively simple model was able to produce reasonable simulations of ET and biomass production over the 27-day period of this study. ET and LAI estimated from NOAA AVHRR observations were effective in calibrating the soil water and vegetation growth components of the model. The raster GIS was effective in organizing spatially distributed data and model executions. Use of the programming language within the GIS results in great flexibility in coordinating data input, model execution, and output of results. This strategy could provide a means of developing useful analytical tools for resource managers.

ACKNOWLEDGEMENT

The authors wish to thank the participants of the MONSOON 90 Experiment for making this data set available for our use. In particular, we thank William P. Kustas (USDA-ARS, Beltsville, MD), David Goodrich (USDA-ARS, Tucson, AZ), James H. Blanford (Univ. of Arizona, Tucson, AZ), William D. Nichols (USGS, Carson City, NV), Mark A. Weltz (USDA-ARS, Tucson, AZ), M. Susan Moran (USDA-ARS, Tucson, AZ), Thomas R. Clarke (USDA-ARS, Phoenix, AZ), C.S.T. Daughtry (USDA-ARS, Beltsville, MD), and Eileen M. Perry (Battelle Pacific Northwest Laboratories) for guidance and the use of data.

DISCLAIMER

Mention of trade names within this document does not imply endorsement by the U.S. Department of Agriculture.

REFERENCES

Charles-Edwards, D.A. (1982) *Physiological Determinants of Crop Growth*. Academic Press, Sydney, Australia.

Daughtry, C.S.T., and Perry, E.M. (1991) Non-destructive estimates of foliage area index at the METNET sites in Walnut Gulch. *MONSOON 90 Data Report*. USDA-ARS, Beltsville, MD.

Doraiswamy, P.C., and Perry, E.M. (1991) The relationship between satellite-based surface temperature and vegetation indices over a semi-arid rangeland watershed. *Preprints, Tenth Conference on Biometeorology and Aerobiology and Special Session on Hydrometeorology*. American Meteorological Society, pp. 226-229.

Jackson, R.D. (1985) Evaluating evapotranspiration at local and regional scales. *Proceedings, IEEE*. 73: 1086-1096.

Kustas, W.P., Goodrich, D.C., Moran, M.S., Bach, L.B., Blanford, J.H., Chehbouni, A., Claassen, H., Clements, W.E., Doraiswamy, P.C., Dubois, P., Clarke, T.R., Daughtry, C.S.T., Gellman, D., Hipps, L.E., Huete, A.R., Humes, K.S., Jackson, T.J., Nichols, W.D., Parry, R., Perry, E.M., Pinker, R.T., Pinter, P.J., Jr., Qi, J., Riggs, A., Schmugge, T.J., Shutko, A.M., Stannard, D.I., Swiatek, E., Vidal, A., Washburne, J., and Weltz, M.A. (1991a) An interdisciplinary field study of the energy and water fluxes in the atmosphere-biosphere system over semiarid rangelands: Description and some preliminary results. *Bulletin of the American Meteorological Society*. 72: 1683-1706.

Kustas, W.P., Perry, E.M., and Doraiswamy, P.C. (1991b) MONSOON 90 - A remote sensing feasibility study of the energy and water balance of a semiarid basin. *42nd Congress of the International Astronautical Federation*. Paper IAF-91-160.

Kustas, W.P., Perry, E.M., Doraiswamy, P.C., and Moran, M.S. (1994) Using satellite remote sensing to extrapolate evapotranspiration estimates in time and space over a semiarid rangeland basin. *Remote Sensing of Environment*. 49: 275-286.

Maas, S.J. (1988a) Use of remotely sensed information in agricultural crop growth models. *Ecological Modelling*. 41: 247-268.

Maas, S.J. (1988b) Using satellite data to improve model estimates of crop yield. *Agronomy Journal*. 80: 655-662.

Maas, S.J. (1991a) Validation of GRAMI wheat yield estimates for North Dakota using Landsat MSS data. *Preprints, 20th Conference on Agricultural and Forest Meteorology*. American Meteorological Society, pp. 228-231.

Maas, S.J. (1991b) Use of remotely sensed information in plant growth simulation models. *Advances in Agronomy*. Council for Scientific Research Integration, Trivandrum, India. 1: 17-26.

Maas, S.J. (1992) *GRAMI: A Crop Growth Model That Can Use Remotely Sensed Information*, ARS-91. U.S. Department of Agriculture, Washington, DC.

Maas, S.J. (1993a) Parameterized model of gramineous crop growth: I. Leaf area and dry mass simulation. *Agronomy Journal*. 85: 348-353.

Maas, S.J. (1993b) Parameterized model of gramineous crop growth: II. Within-season simulation calibration. *Agronomy Journal*. 85: 354-358.

Maas, S.J., Moran, M.S., and Jackson, R.D. (1992) Combining remote sensing and modeling for regional resource monitoring, Part II: A simple model for estimating surface evaporation and biomass production. *Technical Papers, ASPRS/ACSM/RT92 Convention*. American Society for Photogrammetry and Remote Sensing, pp. 225-234.

Maas, S.J., Moran, M.S., Weltz, M.A., and Blanford, J.H. (1993) Model for simulating surface evaporation and biomass production utilizing routine meteorological and remote sensing data. *Annual ACSM/ASPRS Convention*, American Society for Photogrammetry and Remote Sensing, pp. 212-221.

Moran. M.S., and Jackson, R.D. (1991) Assessing the spatial distribution of evapotranspiration using remotely sensed inputs. *Journal of Environmental Quality*. 20: 725-737.

Moran. M.S., Maas, S.J., and Jackson, R.D. (1992) Combining remote sensing and modeling for regional resource monitoring, Part I: Remote evaluation of surface evaporation and biomass. *Technical Papers, ASPRS/ACSM/RT92 Convention*. American Society for Photogrammetry and Remote Sensing, pp. 215-224.

Moran, M.S., Maas, S.J., and Pinter, P.J., Jr. (1995) Combining remote sensing and modeling for estimating surface evaporation and biomass production. *Remote Sensing Reviews*. 12: 335-353.

Van Bavel, C.H.M. (1966) Potential evaporation: The combination concept and its experimental verification. *Water Resources Research*. 2: 455-467.

Stephan J. Maas
Plant Physiologist, USDA-ARS Shafter Research Station
17053 Shafter Avenue
Shafter, California, USA 93263
Telephone: (805) 746-8002
FAX: (805) 746-1619
Email: smaas@academic.csubak.edu

Rusty F. Dodson, David P. Turner

Using GIS to enable diagnostic interaction with a spatially distributed biogeochemistry model

Spatially distributed ecological modeling relies heavily on GIS techniques to prepare input data and to manipulate, validate, and visualize output data. In an on-going project to model climate-change effects on plant productivity and water availability in the Columbia River Basin (670,000 km^2), GIS techniques are used at every stage in the modeling effort. Initially, a point-mode version of a daily time step biogeochemistry model (Forest-BGC) was adapted to a spatially distributed mode by modularizing the source code and embedding it in an Image Processing Workbench (IPW) shell. The climatological inputs to the model (daily surfaces of solar radiation, precipitation, and minimum/maximum/dewpoint temperature) are prepared, checked, and integrated within a GIS framework. Model outputs, such as daily photosynthesis and evapotranspiration, may be saved as multi-band images (one band per day) or single-band annual totals. The manipulation and spatial query tools provided by GIS are used in validating modeled output (e.g. comparing predicted and measured stream discharge) and in sensitivity analysis (e.g. investigating the model's response to different grid resolutions). Visualization and data exploration GIS tools provide insights to the behavior of the model, facilitating further model development and enhancement. To this end, an exploratory space-time analysis environment was created which allows the user to display a map, (e.g. initial vegetation type), click on a grid cell, and run the point-based model at the selected location. An interactive display environment then provides access to daily values for over 60 variables computed by the model. The visualization of both spatially distributed and point-based temporal data provides a powerful tool for model calibration, validation, and refinement.

INTRODUCTION

In modeling the potential effects of climate change on ecosystems, there may be virtue in retaining complexity. Models which are highly aggregated in terms of compartments, and which run at aggregated time steps (e.g. annual), may miss interactions which hold the key to understanding potential ecosystem responses. Because of intense interest in "winners and losers", and because changes along environmental gradients mean that the responses will be unique to each location, it is also desirable to apply these process-based models in a spatially distributed mode. Increasingly, the creation of a relevant modeling framework (one which retains complexity and is spatially distributed) is not constrained by the availability of input data or by computer processing speed, but rather by the tasks of integrating the data inputs and maintaining adequate model diagnostic capability. Discerning causation is a crucial aspect of simulation modeling. In this paper, we describe a modeling framework for use in climate change assessments which makes extensive use of GIS for data management, model validation and model diagnostics. The modeling framework makes it feasible to run a complex process-based ecosystem model at relatively fine spatial and temporal resolutions over large geographic regions.

THE MODEL

The modeling framework discussed in this paper began with the Forest-BGC (Biogeochemical) model (Running and Coughlan 1988), an ecosystem process model which simulates hydrology and carbon flux at a daily time step over a single point in space. The point-based BGC model was translated to the C programming language (Kernighan and Ritchie 1988) and then extended to a spatially distributed mode by embedding it within an IPW (Image Processing Workbench) shell. IPW (Frew 1990, Longley et al 1992) is a public-domain software library which provides a large set of spatial analysis and spatial data management functionality for the image, or raster, data structure. The IPW implementation of the BGC model, IBGC, takes a set of images as input, runs the point-based BGC model at each image pixel, and outputs a set of images which correspond to the output variables requested by the user.

Because it provides no display or mapping functions, IPW is not a GIS *per se*, based on a conventional definition of GIS as a system for the input, manipulation, analysis, and display of spatial data (Goodchild 1994). However, IPW does provide GIS capabilities which are invaluable in spatially distributed modeling and which are not available in any commercial GIS packages at present (e.g. the ability to handle and analyze 356-band images, the computation of solar radiation over a topographic surface). To fill the missing GIS component--spatial data display--a link was built between IPW and PV-WAVE (Visual Numerics Inc. 1992), which allows for "quick and dirty" visualization of maps, histograms, and spatial transects. A further link was built between PV-WAVE and the IBGC model itself, which allows full interaction between the user and the model. The integrated modeling framework is depicted in Figure 1. While viewing an on-screen image (e.g. vegetation type), the user selects a location and the model is run at that site. A visualization environment is then presented which allows a time- series display of virtually every variable within the model (over 75 variables).

[Figure 1 - The integrated spatial modeling framework.]

MODEL INPUTS AND GIS

The IBGC model is driven by two types of input data: aspatial, and spatial. The aspatial inputs consist of simple text files of constants (e.g. maximum stomatal conductance for each biome type) and model parameters (e.g. simulation duration, requested output variables). For its spatial input, IBGC requires daily surfaces of precipitation, solar radiation, and minimum, maximum, and dewpoint temperature. These surfaces are stored and managed as 365-band IPW images, where each image band contains one day's worth of data. An additional image-based input is the 23-band *parameter image*, which includes temporally constant but spatially varying parameters such as soil carbon, soil water holding capacity, and biome type (currently conifer, grass, or shrub). GIS techniques are used extensively in the production of the climate inputs. With the exception of the monthly precipitation source images, all of the methods described below to create climatic inputs to the IBGC model are implemented entirely within IPW. Figure 2 shows maps of several types of input surfaces used in a study which modeled the Willamette/Deschutes river basins (approximately 55,000 km^2) in Oregon at a 1-km grid resolution.

[Figure 2 - Map of temperature, precipitation, radiation, LAI, and soil WHC for the Willamette/Deschutes river basins in Oregon.]

Air Temperature

For minimum/maximum air temperature, a set of daily meteorological station observations were converted to 1000 mb (approximately sea-level) potential temperatures, interpolated by an inverse-distance weighting algorithm, and the resulting potential temperature surfaces adjusted to actual temperature based on elevation data represented by a digital elevation model (DEM) (Marks 1990). The resulting temperature surfaces conformed to the horizontal temperature patterns represented by the meteorological station observations and also to the vertical temperature patterns expected over the DEM terrain under neutrally-stable atmospheric conditions.

For the dewpoint temperature input, observations of dewpoint, if available at sufficient spatial resolution, were converted to relative humidities, distributed over DEM terrain, and converted back to dewpoint temperatures (Marks 1990) using modeled mean temperature surfaces. In the absence of detailed dewpoint data, minimum air temperature was used as a surrogate. This has been shown to be a reasonable assumption in a mountainous region of the western U.S. (Glassy and Running, 1994).

Precipitation

For the precipitation input, the starting data consisted of a set of monthly precipitation surfaces created by the PRISM model (Daly et al 1994). The PRISM model interpolates precipitation station measurements over complex terrain through a moving-window local regression technique which estimates precipitation as a function of elevation for a set of neighboring stations on similar topographic exposures. To derive daily precipitation surfaces from the monthly PRISM surfaces, a set of daily precipitation measurements was required. For each month, daily meteorological station data were used to interpolate (with an inverse-distance algorithm) a series of daily proportion surfaces, where each surface represents the proportion of a month's precipitation that fell on a given day (Daly 1994). The proportion surfaces were normalized to ensure that each month sums to 1.0. The daily proportion surfaces were then used to modulate the monthly PRISM surface to make daily precipitation surfaces.

Solar Radiation

The radiation surfaces were computed by a multi-step process. First, for 20 significant days throughout the year (e.g. the solstices), a clear-sky radiation surface was created using a topographic radiation model which computes direct-beam, diffuse, and terrain-reflected irradiance over a DEM (Dozier 1980, Dubayah et al 1990). Next, daily clear-sky radiation surfaces were linearly interpolated between each of the 20 significant days. Finally, cloudiness coefficients for each day were computed as a function of diurnal temperature range (Bristow and Campbell 1984) and used to modulate the daily clear-sky radiation surfaces.

Land Cover and Leaf Area Index

Besides the climate data, model inputs include the biome type and leaf area index (LAI). In recent years, the availability of digital land cover maps based on satellite imagery has greatly increased (e.g. Loveland et al. 1991). These maps provided the basis for the biome-level land cover surfaces used in IBGC. Information on the forest age-class distribution is also desirable for the purposes of initializing variables such as stem carbon and litter carbon. Remote sensing has been applied in age class discrimination of coniferous and deciduous forests (Hall et al. 1991, Cohen et al. In Press) but this information has not yet been employed within IBGC

Remote sensing also offers promise for estimating LAI (Badhwar and MacDonald 1986, Spanner et al. 1994). However, empirical relationships of LAI to vegetation indices such as NDVI tend to be asymptotic at high LAIs (Peterson et al. 1987) and may be sensitive to terrain slope and aspect. Alternatively, potential LAI may be estimated from regressions between observed LAI and a water balance index derived from distributed meteorological data (Grier and Running 1977). In coniferous forests of the northwestern U.S., an LAI surface for input to IBGC was created (Turner et al. In Press) based on a regression of LAI against a water balance index (sum of growing season precipitation plus soil water holding capacity minus growing season potential evapotranspiration).

Soil Water Holding Capacity

Soil variables such as water holding capacity and carbon pool size are also needed for model initialization. Soils databases with associated spatial data, such as NATSCO and STATSCO (Lytle 1993), have been used to develop digital maps of these properties (e.g. Kern 1995) but the map units may be large relative to the spatial resolution of climate and land cover data. Higher spatial resolution soils data may exist in some cases and high resolution DEMs have been used to refine map unit averages (Nemani et al. 1993). We used the STATSCO-based data from Kern 1995.

Once all the input data are generated, GIS techniques are required to make the various images compatible. Some of the climate variables (IPW solar radiation and PRISM precipitation) were created within a latitude/longitude coordinate system. Since the IBGC model is often run in a different map projection (e.g. Albers equal-area conic), the latitude/longitude data were projected into the desired model projection. The input images also had to be co-registered in space to the same grid resolution, which required the grid resampling and smoothing capabilities of a GIS.

MODEL OUTPUTS AND GIS

In addition to facilitating the creation and management of input data, GIS techniques are useful in evaluating and analyzing the model's output data. The IBGC model outputs requested variables as images, which can be written out daily, monthly, or annually. Typical output images include daily evapotranspiration (ET), daily soil decomposition rate, monthly stream discharge, and annual net primary productivity (NPP). The first step in interpreting the IBGC output was to attempt to validate it through comparison with measured data.

Validation of Stream Discharge

The IBGC model has typically been run over a region containing one or more large river basins (e.g. the 55,000 km^2 Willamette/Deschutes river basins of Oregon, or the 830,000 km^2 Columbia River Basin which comprises most of Oregon, Washington, and Idaho). For most of the potential output variables, such as net primary productivity (NPP) or evapotranspiration (ET), there are no reliable sets of observational data which are available over the entire study region. One exception to this is stream discharge, which is meticulously recorded and quality-checked by the United States Geological Survey (USGS) over a relatively dense network of measurement stations throughout the U.S. (USGS 1993). While stream discharge is measured at a single point in space, each measurement corresponds and is sensitive to the hydrologic properties of its entire upstream watershed. Thus a small number of strategically-located discharge measurements can effectively represent the hydrologic activity of a large area.

Of course, discharge data are not free from uncertainty. Irrigation diversions can remove significant amounts of water from the expected streamflow, and dam-regulated flow can seriously alter the seasonal discharge pattern by decreasing the outflow from very high-flow events and releasing stored water during low-flow periods. Since the USGS provides notes on irrigation and flow regulation for each measurement site, careful selection of watersheds can often minimize these errors.

A multi-step GIS analysis was required to convert IBGC-modeled discharge values into values which were directly comparable with USGS measurements. First, for each measurement site, a watershed area was delineated. This procedure can be performed within standard GIS packages such as ARC/INFO and GRASS, and requires the following steps: filling any sinks or depressions within the DEM, computing the flow directions and accumulations for each DEM grid cell, snapping the measurement site location to the nearest grid cell of the DEM-defined stream network (the watershed *pour point*), and then identifying all cells upstream from the pour point. Next, the area of the DEM-derived watershed was compared to the watershed area reported by the USGS. If the two areas were drastically different from each other (we used a threshold of 10 percent) then that watershed was considered problematic and was discarded. Possible sources of these areal discrepancies include inability of the spatial scale of the DEM to represent the true basin topography (White and Running 1994), improperly-snapped pour point locations, and uncertainty in the locational and areal data reported by the USGS.

If the basin areas (observed and modeled) were comparable, then observed and modeled monthly discharge values were converted to units of mm in order to normalize any remaining basin area discrepancies (i.e. within the 10% difference threshold). Monthly discharge hydrographs were then plotted and analyzed for consistency in terms of both total flow amounts and seasonal flow patterns. Figure 3 shows a monthly hydrograph of observed and modeled discharge at the mouth of Oregon's Willamette River (basin area: 28,700 km^2; modeled grid resolution: 1-km). The January and April overshoots in the modeled data are probably due to the simple bucket-style soil hydrology used by the Forest-BGC model, which tends to produce a faster and more pronounced response to precipitation events than more complex subsurface-flow-based soil hydrologies (Nemani et al. 1993). Despite the discrepancies in seasonal pattern, the total annual discharge values were in agreement to within 4% (observed = 833 mm, modeled = 803 mm).

[Figure 3 - Monthly modeled and observed discharge: Willamette River, Oregon, water year 1990.]

Validation of Carbon Flux

For carbon flux validation, there is no analog to streamflow, i.e. no measurable parameter which integrates flux measurements over a large area. Thus, point-based measurements over the environmental gradients within the modeled domain are necessary. At the daily time step, Forest-BGC estimates photosynthesis and plant respiration, and we have modified it to estimate the daily CO2 efflux from decomposition of soil, litter and woody debris. The model therefore produces estimates of net 24 hour CO2 flux which can be compared with eddy correlation flux measurements. Model estimates of soil CO2 efflux, the sum of root respiration and heterotrophic respiration, can likewise be compared with chamber-based measurements (e.g. Mattson 1995). We have not yet performed these carbon flux validations.

At an annual time step, measurements of net primary production based on litterfall and basal area increments have been used to validate Forest-BGC estimates (Running 1994). A great deal of forest productivity studies have been made because of the economic significance of forest growth, but these studies are rarely georeferenced and, because of their economic significance, may be proprietary. Large, interdisciplinary field campaigns (Peterson and Waring 1994), in which a complete suite of carbon flux measurements is made, are ideal for calibration and validation purposes.

Sensitivity Analysis

Another key role played by GIS in our modeling framework is in assessing the sensitivity of the model results to various parameters or initial conditions. To analyze IBGC's sensitivity to the spatial resolution of the grid, the model was run over the Willamette and Deschutes watersheds (approximately 55,000 km^2) in Oregon at three different grid resolutions: 1-km, 10-km, and 50-km. GIS techniques were essential in appropriately resampling the 1-km data to the coarser resolutions (e.g. categorical data such as biome type require different resampling techniques than do continuous data). GIS also played an important role in interpreting and combining the outputs from the three model runs in meaningful ways. Differences in modeled NPP from the 1-km and the 50-km model runs indicate the loss of spatial information incurred at the coarser resolution (Figure 4).

[Figure 4 - Absolute difference in NPP modeled at 1-km and 50-km resolution, Willamette/Deschutes basins.]

DIAGNOSTIC INTERACTION

IBGC has the potential to output a huge amount of data. For a moderately-sized study area of 500x500 grid cells, a daily output file for a single variable takes up over 90 megabytes of disk space. There are over 50 variables that can be output on a daily basis. Once the model has been properly calibrated, a monthly or annual aggregation is usually sufficient for analyzing and presenting the model output. However, the daily time step is often of crucial interest at the early modeling stages where initial parameters are being calibrated, and also when analyzing the model's behavior in order to identify areas for future enhancement.

Instead of sifting through vast volumes of daily image data, the IBGC model can be run interactively in point mode through the IBGC/PV-WAVE interface. This gives the investigator full access to all internal IBGC variables at the daily timestep. Also, since the point-mode model may be run for any pixel in the study area, the investigator can explore the model's behavior over space by selecting pixels along any desired environmental gradient.

Figure 5 shows a session of the IBGC/PV-WAVE interface in action. The image at the upper-right is a shaded-relief map of a 2.5-arc-minute (approximately 4-km) DEM of Washington, Oregon, and Idaho. The area on the image labeled "point 1" was selected with the mouse, and designates a forested site in the Oregon Coast Range. The window at the upper-right of Figure 5 is a selection menu of daily-time-step variables. The user may select up to 4 variables to be plotted on the left-side Y-axis and 4 for the right-side Y-axis. The lower window of Figure 5 contains the results of a one-year run of the point-based IBGC model at the selected site with the following daily selected variables: soil water fraction (proportion of capacity), soil decomposition rate (proportion of maximum potential), cumulative net carbon (kg/ha), and cumulative autotrophic respiration (kg/ha). The left Y-axis corresponds to the 0-1 proportion variables, the right Y-axis corresponds to the kg/ha variables, and the X-axis corresponds to the day of the water year (Oct. 1 to Sept. 30). The plot window of Figure 5 shows several patterns: soil water depletion, soil decomposition, and slight carbon accumulation during days 0-50 (Oct. 1 - Nov. 20), followed by soil water at capacity and minimal soil decomposition during the cold wet period of days 50-160 (Nov. 17 - Mar. 7). During late spring and early summer net carbon accumulates rapidly while the soil steadily dries out, interrupted by an occasional storm. Carbon accumulation drops off when the soil water gets very low (at around day 310), and picks up again after some September rainfall.

[Figure 5 - The IBGC/PV-WAVE interface: one site, several variables.]

Figure 6 is from the same modeling session depicted in Figure 5, but shows a single variable (net carbon accumulation) over three sites in Washington's Olympic Peninsula. In the image window of Figure 6, the three sites (points 1, 2, and 3) lie along a gradient of decreasing elevation on the south slope of Mount Olympus. The corresponding plot of net carbon shows that the highest elevation site (point 1) is least productive, and the lowest site (point 3) is most productive. The plot also nicely depicts the relationship between elevation (and thus temperature) and the length of the growing season.

[Figure 6 - The IBGC/PV-WAVE interface: three sites, one variable.]

Visualizations like the ones depicted in Figs. 5 and 6 provide powerful exploratory insights to the relationships among variables within the model and across space within the study region. The IBGC/PV-WAVE interface played a key role in detecting and evaluating a bias in an earlier version of the air-temperature interpolation algorithm by providing views of temperature and modeled snowpack for high-elevation sites. The detection of the temperature bias launched a separate study on temperature interpolation that produced a modified algorithm which greatly reduced the bias.

FUTURE GIS ISSUES

Spatially distributed ecosystem modeling at a daily time step presents a unique set of challenges. First, the development of daily climate surfaces requires detailed and often complex interpolation procedures. Many of these procedures are relatively new and few of them exist within commercially-available GISs. Second, the handling of 365 daily surfaces per climate variable per year is extremely difficult with the crude or nonexistent multi-band image tools offered by standard GIS software. An important aspect of our modeling framework is the ability to easily create and manipulate 365-band files of floating-point image data through the use of IPW. Each IBGC model run requires 1,848 individual images, which are stored as 6 IPW multi-band image files (minimum, maximum, dewpoint temperature, radiation, precipitation, and the 23-band parameter image).

Considerable flexibility is gained by linking together several "lean" systems--systems which operate on simple, open data formats; which use a straightforward, often command-line-based user interface; and which don't attempt to provide "bells and whistles" for every possible application. IPW is one such lean system which has two key features for integration with models. First, access to the source code allows a model to be incorporated directly within the system. Source code access also permits the construction of custom spatial analysis tools such as those used to derive daily precipitation surfaces from daily measurement points and monthly surfaces. Second, IPW image files are simple, low-level files which are easily passed to other applications. These features make IPW an extremely powerful spatial analysis and modeling engine. PV-WAVE, used as the display engine, is also a lean system as it easily reads flat binary image data and provides very quick pop-up graphics windows which don't require the overhead of starting up a large, complicated software package.

The future in integrating GIS and environmental models lies in either linking together lean systems, or in embedding models within full-function GIS packages. Both methods are promising, yet each has its difficulties. Lean system linkage requires the knowledge and use of several software systems, each with a different user-interface, scripting language, etc. Integration with full-function GIS is impeded by complex, often proprietary file formats. As the GIS user base becomes more sophisticated, future demand for integrated modeling may force GIS vendors to open up their systems. At the present time, however, lean system linkage seems to be the most feasible and flexible option.

CONCLUSION

The issue of "scaling up" from point based observations and model simulations of biogeochemical processes to basin-wide or regional flux estimates has become increasingly important. Climate change analyses, in particular, require spatially distributed models because of the continuous nature of climatic variation in space. Integration of complex simulation models and spatial data is necessary and can be greatly facilitated by existing and emerging GIS tools. Essential features of a viable spatially distributed modeling framework include the ability to rapidly visualize variables as surfaces at particular time slices, or plotted over time for a specified point. The modeling framework presented here achieves that capability by linking a point-based model, an image processing shell, and a data visualization tool.

CLEARANCE

The information in this document has been funded (wholly or in part) by the U.S. Environmental Protection Agency under Contract 68-C4-0019 to Mantech Environmental Research Services Corp. It has been subjected to Agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ACKNOWLEDGEMENTS

We thank Steven Running for providing the source code to Forest-BGC, and Dana Jacobsen and Greg Koerper for their programming work on earlier versions of the IPW and PV-WAVE model components.

REFERENCES

Badhwar G. D., and R. B. MacDonald. (1986) Satellite-derived leaf-area-index and vegetation maps as input to global carbon cycle models-a hierarchical approach. International Journal of Remote Sensing 7:265-281.

Bristow, K. L., and Campbell, G. S. (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agricultural and Forest Meteorology 31:159-166.

Cohen, W.B., T. A. Spies, and M. Fiorella. (1995) Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A. International Journal of Remote Sensing 16:721-746.

Daly, C. (1994) Modeling climate, vegetation, and water balance at landscape to regional scales. PhD dissertation, Oregon State University.

Daly, C., Neilson, R. P., and Phillips, D. L. (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. Journal of Applied Meteorology 33:140-158.

Dozier, J. (1980) A clear-sky spectral solar radiation model for snow-covered mountainous terrain. Water Resources Research 16:709-718.

Dubayah, R., Dozier, J., and Davis, F. (1990) Topographic distribution of clear-sky radiation over the Konza Prairie, Kansas. Water Resources Research 26:679-690.

Frew, J. S. (1990) The image processing workbench. Ph.D. dissertation. Department of Geography, University of California, Santa Barbara, California, 382 pp.

Glassy, J.M., and Running, S.W. (1994) Validating diurnal climatology logic of the MT-CLIM model across a climatic gradient in Oregon. Ecological Applications 4:248-257.

Goodchild, M. F. (1994) Integrating GIS and remote sensing for vegetation analysis and modeling: methodological issues. Journal of Vegetation Science 5:615-626.

Grier, C. C., and S. W. Running. (1977) Leaf area of mature northwestern coniferous forests: Relation to site water balance. Ecology 58:893-899.

Hall, F.G., D.B. Botkin, D. E. Strebel, K.D. Woods, and S.J. Goetz. (1991) Large-scale patterns of forest succession as determined by remote sensing. Ecology 72:628-640.

Kern, J.S. (1995) Geographic patterns of soil water holding capacity in the contiguous United States. Soil Science Society of America Journal 59:1126-1133.

Kernighan, B. W., and Ritchie, D. M. (1988) The C Programming Language. New Jersey: Prentice Hall, 272 pp.

Longley, K. D., Jacobsen, D., and Marks, D. (1992) Supplement to the Image Processing Workbench (IPW): modifications, procedures, and software additions, Revision 2.0. Technical Report, EPA-COR, Corvallis, OR: U.S. EPA Environmental Research Laboratory.

Loveland, T. R., J. W. Merchant, D. O. Ohlen, and J. F. Brown. (1991) Development of a land-cover characteristics database for the conterminous US. Photogrametric Engineering and Remote Sensing 57:1453-1463.

Lytle, D. J. (1993) Digital soils databases for the United States. Chap. 38, pages 386-391 in: Goodchild, M.F., Parks, B.O., and Steyaert, L.T. (eds.) Environmental modeling with GIS, New York, NY: Oxford University Press.

Marks, D. (1990) A continental-scale simulation of potential evapotranspiration for historical and projected doubled-CO2 climate conditions. In: Guchinski, H., Marks, D., and Turner, D. (eds) Biospheric Feedbacks to climate Change: The Sensitivity of Regional Trace Gas Emissions, Evapotranspiration, and Energy Balance to Vegetation Redistribution, U.S. Environmental Protection Agency report EPA/600/3-90/78, Corvallis, OR, pp. III/1 - III/44.

Mattson, K.G. (1995) CO2 efflux from coniferous forest soils: comparison of measurement methods and effects of added nitrogen. In: Lal, R., Kimble, J.M., and Levine, E., (eds.) Soil Processes and Greenhouse Effect. London: Lewis Publishers.

Nemani, R., S. W. Running, L. E. Band, and D. L. Peterson. (1993) Regional hydroecological simulation system: An illustration of the integration of ecosystem models in GIS. Chap. 28, pages 296-304 in: Goodchild, M.F., Parks, B.O., and Steyaert, L.T. (eds.) Environmental modeling with GIS, New York, NY: Oxford University Press.

Peterson, D. L., and R. H. Waring. (1994) Overview of the Oregon transect ecosystem research project. Ecological Applications 4:211-225.

Peterson, D. L., M. A. Spanner, S. W. Running, and K. B. Teuber. (1987) Relationship of thematic mapper simulator data to leaf area index of temperate coniferous forests. Rem. Sens. of the Envir. 22:323-341.

Running, S.W. (1994) Testing forest-BGC ecosystem process simulations across a climatic gradient in Oregon. Ecological Applications 4:238-247.

Running, S. W. and J. G. Coughlan (1988) A general model of forest ecosystem processes for regional application. Ecological Modelling 42:125-154.

Spanner, M., L. Johnson, J. Miller, J. Freemantle, J. Runyon, and P. Gong. (1994) Remote sensing of seasonal leaf area index across the Oregon transect. Ecological Applications 4:258-271.

Turner, D. P., Dodson, R. F., and Marks, D. (1995) Comparison of alternative spatial resolutions in the application of a spatially distributed biogeochemical model over complex terrain. Ecological Modelling. In press.

U.S. Geological Survey (1993) Hydro-Climatic Data Network (HCDN): Streamflow data set, 1874-1988. U.S. Geological Survey water-resources investigations report 93-4076, Washington, D.C: U.S. Department of the Interior, CD-ROM.

Visual Numerics Inc. (1992) PV-WAVE Command Language Reference. Boulder, CO: Visual Numerics, Inc.

White, J.D., and Running, S.W. (1994) Testing scale dependent assumptions in regional ecosystem simulations. Journal of Vegetation Science 5:687-702.

Rusty F. Dodson
ManTech Environmental Research Services Corp.
National Health and Environmental Effects Research Laboratory
Western Ecology Division
200 S.W. 35th Street
Corvallis, OR 97330
Phone: 503/754-4637
FAX: 503/754-4818
Email: dodson@mail.cor.epa.gov

David P. Turner
ManTech Environmental Research Services Corp.
National Health and Environmental Effects Research Laboratory
Western Ecology Division
200 S.W. 35th Street

Corvallis, OR 97330
Phone: 503/754-4769
FAX: 503/754-4818
Email: dave@mail.cor.epa.gov

Dr. John Yarie

A Forest Ecosystem Dynamics Model Integrated within a GIS.

ABSTRACT

Modeling the biology of forest ecosystems has been devoted to a combination of theoretical and empirical approaches representing the function of a forest ecosystem generally within an undefined spatial context. Moving to a large spatial context will require the use of theoretical representations of critical ecosystem functions that can be represented on an individual cell basis. It should then be possible to vary the size of the smallest cell from 1 m2 to 100 ha.

A forest ecosystem dynamics model is being developed that is based on the nitrogen productivity concept for forest growth; litterfall quality and microbial efficiency for forest floor decomposition, and forest regeneration based on a tree's sprouting or seed production capability. Climate and ecosystem level disturbances will be handled as restricted stochastic processes. The restriction will be based on known state factor relationships. The state factors are used to describe a broad scale classification of the landscape to define basic limitations for the randomly derived driving variables used in the model.

The model has been programed as an ARC/INFO AML within the GRID package. The current version of the model has been verified as functional from an individual tree basis (1 m2 cell size) in a number of forest types found in interior Alaska. Verification on a landscape scale (1 ha. cell size) is difficult because of a lack of detailed data that can be used from a landscape perspective.

INTRODUCTION

Calculation of carbon source/sink relationships for large land areas are often based on a simple calculation of the dominant vegetation usually at a mature state for the land area. In some instances the landscape is represented by a relatively simple classification of landscape characteristics (e.g. coniferous vs. deciduous forest). In these types of calculations, differences in carbon dynamics due to potentially important factors; like topography, soils, differences in climate, or variation in vegetation community types, and the community age structure; within a region are not considered. The calculation would only be accurate if an appropriate landscape weighted average was chosen for the region's carbon factor.

A geographically referenced group of data sets that can be used to define the primary state factors that control ecosystem function across Alaska is being developed. Topography is one of the easiest and can be developed from the DEM data that is currently available. A climate classification has been developed (Hammond and Yarie, in press). A preliminary version of the land characteristics for the state has been developed (Flemming, pers comm), and work is currently in progress to develop an age structure map of the forest types within the state.

These data sets will then be used to parameterize a Geographic Alaskan Forest Ecosystem Dynamics model (GAFED) and to define specific climatic changes in the defined regions that are likely to occur due to global change.

The model will be designed to work at all levels of spatial resolution which is one advantage of incorporating it into ARC/INFO. Primary analysis will be developed at the individual tree within a stand level of landscape resolution (one square meter grid cell resolution). The biogeoclimatic classification for the state of Alaska will then be used to summarize stand level results at the landscape level (one hectare or greater grid cell resolution).

The primary milestone at completion of this work will be the development of a carbon balance map for the state of Alaska with current vegetation and average climate conditions. Changes in carbon dynamics can be estimated based on climate scenarios developed from mesoscale climate models. The biogeoclimatic classification should give us the ability to describe the appropriate level of landscape summarization for Alaska.

MODEL CONCEPTUAL BASIS

The GAFED model is primarily a process model that will use the important limiting factors to drive forest growth, forest floor, and mineral soil dynamic routines. The model was developed as an AML within ARC/INFO GRID. The routines necessary for the model can be developed so that the grid cell size is not a limitation (Table 1). The majority of routines can be applied at either an individual tree (1 m grid cell size) or landscape representation (1 ha or above grid cell size). There are a small group of routines that also require a greater level of modeling detail if used at a small grid cell size (1 m) (Table 1).

Table 1: RELATIONSHIP BETWEEN MODEL ROUTINES AND GRID CELL SIZE.

Routines valid across all grid cell sizes	**Routines that have cell size dependencies**
Production	Litterfall
Decomposition	Regeneration
Climate	Single Tree Mortality
Disturbance by Fire	

VEGETATION PRODUCTION

The nitrogen productivity concept (gren 1983, 1985; Ingestad 1977, 1980, 1981) is used to model the tree growth at both the individual tree and forest stand level (Yarie, in review). The nitrogen productivity can be defined as the amount of annual production per unit of foliar nitrogen:

$$dW/dt = Pn * N \qquad (1)$$

where W is plant or stand biomass, t is time, Pn is the nitrogen productivity (unit production/unit nitrogen), and N is the foliar nitrogen content

At steady state nutrition (d(N/W)/dt = 0) the plants (or forest stands) growth rate is proportional to the amount of foliar nitrogen in the plant (N) and the nitrogen productivity (Pn). The nitrogen productivity is at a maximum during the exponential growth phase and depends on a number of plant properties including genotypic properties, weather conditions, self-shading and ageing. There is a decrease in the nitrogen productivity due to self shading and plant ageing (Ågren 1983) such that:

$$Pn = Pnmax - b * W \tag{2}$$

where Pnmax is the maximum nitrogen productivity, b is considered an ageing and/or light extinction parameter, the other parameters have been defined for equation 1

Equation 2 has been used to calculate the nitrogen productivity of individual seedlings (Ingestad 1979a, 1979b; Ingestad and Kahr 1985) and stands of trees (Ågren 1983). It is also being used to calculate productivity of trees and stands within interior Alaska (Yarie, in review). In both equation 1 and 2 the parameters are not developed for specific geographic unit sizes. It should then be possible to develop a simple equation (equation 2) for calculation of the nitrogen productivity for a single tree to a stand of trees (Yarie, in review).

Development of individual tree and stand level nitrogen productivity curve

The nitrogen productivity of individual trees within a stand was calculated using the 1989 tree chemistry (Yarie and Van Cleve 1996) and above-ground production dataset from the Bonanza Creek LTER program . A total of 239 white spruce, 21 aspen, 54 birch, and 107 balsam poplar trees were available. Because I was trying to estimate the maximum N-productivity for individual trees 37 white spruce, 12 aspen, 8 birch, and 15 balsam poplar were selected for analysis. Individual tree N-productivity was then calculated by dividing the above-ground production by the above-ground foliar nitrogen content.

The comparative analysis between trees and stands was handled by placing all estimates of N-productivity and foliar nitrogen content on a simple unit area basis. The space occupancy of each individual tree was based on a calculation of tree density of a fully stocked stand if the diameter of the sample tree was the average diameter of the stand. The chemical analysis of the foliar material was performed as described by Yarie and Van Cleve (1996).

Calculation of the nitrogen productivity of stands of trees was based on data sets from Van Cleve et al. (1983) and the USFS Inventory of the Porcupine River Drainage (Setzer 1987, Yarie 1983). None of these stands contained any of the trees used for the individual tree calculations. The stands represented independent measurements of the nitrogen content and nitrogen productivity. The foliage quantity per unit area for each stand was again reduced to a one meter square basis.

LITTERFALL

Foliage, root, and twig litterfall is spread equally within a 81, 81 or 121 m2, respectively, area around the tree for the 1 m2 grid cell size. Tree death and stemwood litterfall is positioned in a random direction chosen from eight (0°, 45°, 90°, 315°) potential angles from the tree base. Tree length is calculated based on standard allometric equations relating tree height to tree diameter. In stands of trees with individual grid cell sizes larger than the height of the tallest tree, litterfall occurs within the grid cell.

DECOMPOSITION

The decomposition dynamics are modeled using the theoretical representation presented by Bosatta and Agren (1985) and Agren and Bosatta (1987). Simply, litter quality is set for the fresh litterfall after which it changes by:

(3)

$$\frac{dq}{dt} = -\varepsilon f_c u(q)$$

where E = a scale factor, fc = % C in microbial biomass, and u(q) = microbial growth rate estimated through equation (4)

Microbial growth rate is:

(4)

$$u(q) = u_0 q^{\beta}$$

where u0 = microbial growth rate parameter, q = carbon quality (equation 3), B = growth parameter

Carbon content of the litter cohort decreases at the rate:

(5)

$$\frac{dC}{dt} = \frac{e(q) - 1}{e(q)} f_c C u(q)$$

where e(q) = microbial efficiency, and C = carbon quantity in g/m2

Nitrogen content of the litter cohort decreases at the rate:

(6)

$$\frac{dN}{dt} = \left(\frac{f_N}{f_C} C - \frac{N}{e(q)}\right) f_c u(q)$$

where N = nitrogen quantity (g/m2), fn = microbial % N

Microbial efficiency is:

$$e(q) = e_0 + e_1 q \tag{7}$$

where e0 and e1 are simple coefficients

Model validation will be carried out by using tree growth, forest floor and mineral soil dynamic variables that have been measured in the Fairbanks area as part of the Bonanza Creek Long-Term Ecological Research (LTER) Program (see the BNZ-LTER World Wide Web home page). There is sufficient information available on tree growth and forest floor dynamics from the Bonanza Creek LTER site to evaluate the model behavior for soil temperature, moisture dynamics, carbon and nitrogen turnover, and tree growth across both upland and floodplain successional sequences.

DEVELOPMENT OF A STATE FACTOR BIOGEOCLIMATIC CLASSIFICATION

Currently a number of data sets are available for the state of Alaska. A 90 m DEM derived from USGS data sources is available. This data set can be used to derive relevant elevation, slope and aspect groupings. This data set has been summarized to 1000m grid cell size (USGS EROS field office Anchorage) and was used to develop a topography coverage for Alaska.

The average climatic zones based on a May through September growing season have been determined using average monthly data sets available from NOAA (world wide web home page; http://www.ncdc.noaa.gov) these data sets were brought up to date and edited using locally available weather records. A total of 40 growing season eco-climatic regions and 35 annual eco-climatic regions were estimated for the state of Alaska (Hammond and Yarie, in press).

Work is close to completion on assembling all current vegetation data bases for the state. We will try to summarize this data set at the level IV groups for the Alaska Vegetation Classification (Viereck et al. 1992). The level IV category can be further summarized to broader groupings (e.g. needle-leaf forest, broadleaf forest, etc.) for comparison to datasets derived from other areas of the country. The level IV groupings will also give us the best approach for defining the four primary components of the carbon stores on the landscape. These components are; alive and dead trees (above- and belowground), forest floor, mineral soil, and understory vegetation.

RESULTS

The nitrogen productivity concept represents one approach for developlment of an algorithms for expansion from individual tree to stand or landscape levels of estimation of primary production. A simple nitrogen productivity equation for trees and stands of trees on a unit area basis within interior Alaska was estimated (Figure 1, Yarie, in review).

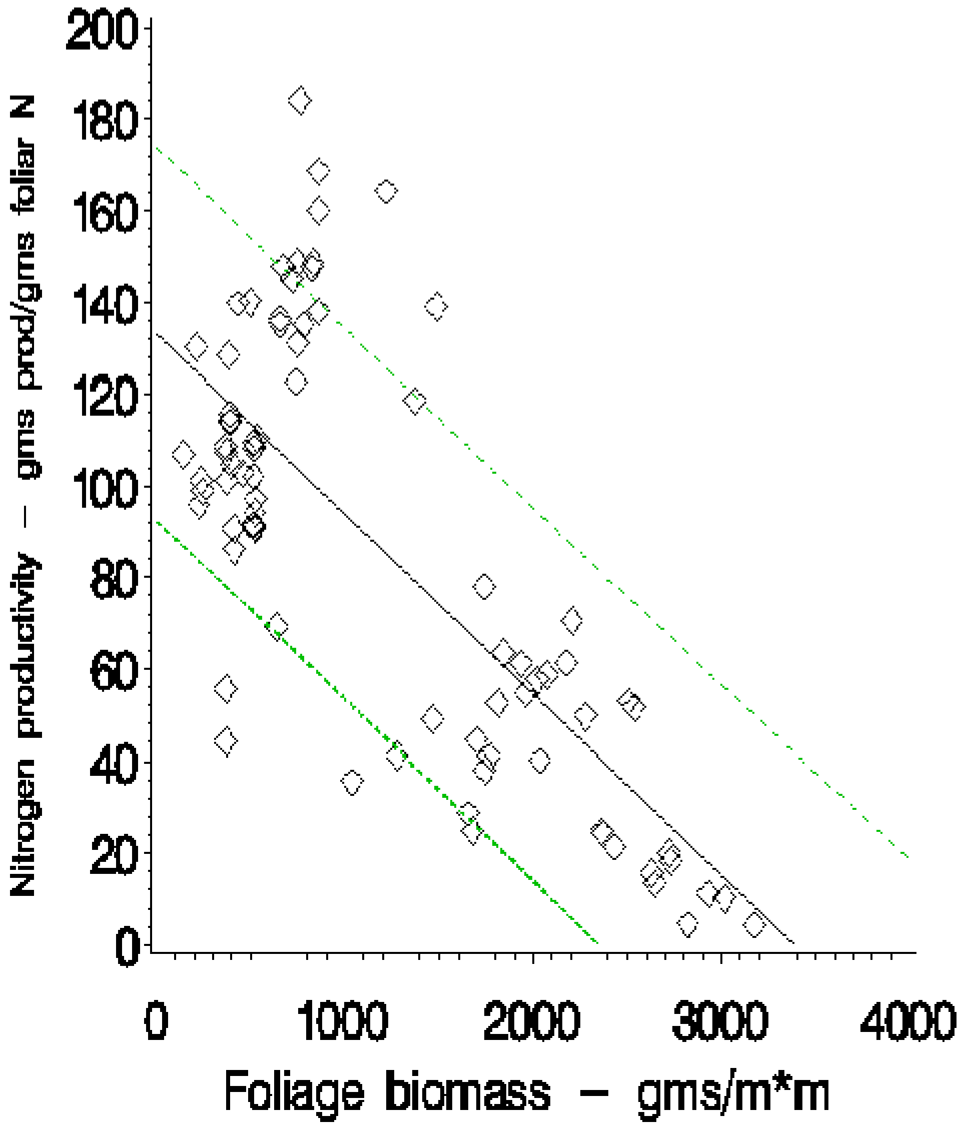


Figure1. Relationship between nitrogen productivity and foliage biomass for aspen, birch, balsam poplar and white spruce stands and individual trees in interior Alaska. The equation is: Maximum N-Productivity = 133.191 - 0.0394 * (foliage biomass/unit area).

The model was able to acurately predict the growth of white spruce and birch trees in an old-growth white spruce forest on the floodplain in interior Alaska. Measured diameter growth for white spruce between 1989 and 1993 averaged 0.9 cm at breast heaight. The model predicted an average of 1.1 cm for the same time period. Total biomass growth for the modeled tree species in this site was approximately 270 gms/m*m. The above ground portion was then approximately 135 gms/m*m. these values are typical for mature white spruce forest stands found in interior Alaska (Yarie and Van Cleve 1983).

The model was able to predict litter decomposition for the tree foliage found on the site when compared to litterbag decomposition from the mature white spruce site (Figure 2). In both cases, the model and litterbag data, were a mixture of the tree foliage litterfall found on the validation site. The exact mixture of the litterfall in the modeled version was dependent on the movement of foliage litterfall around the tree. The model also calculated the decomposition of moss litter, tree twig litter and stemwood litter if a large tree dies and falls to the ground.

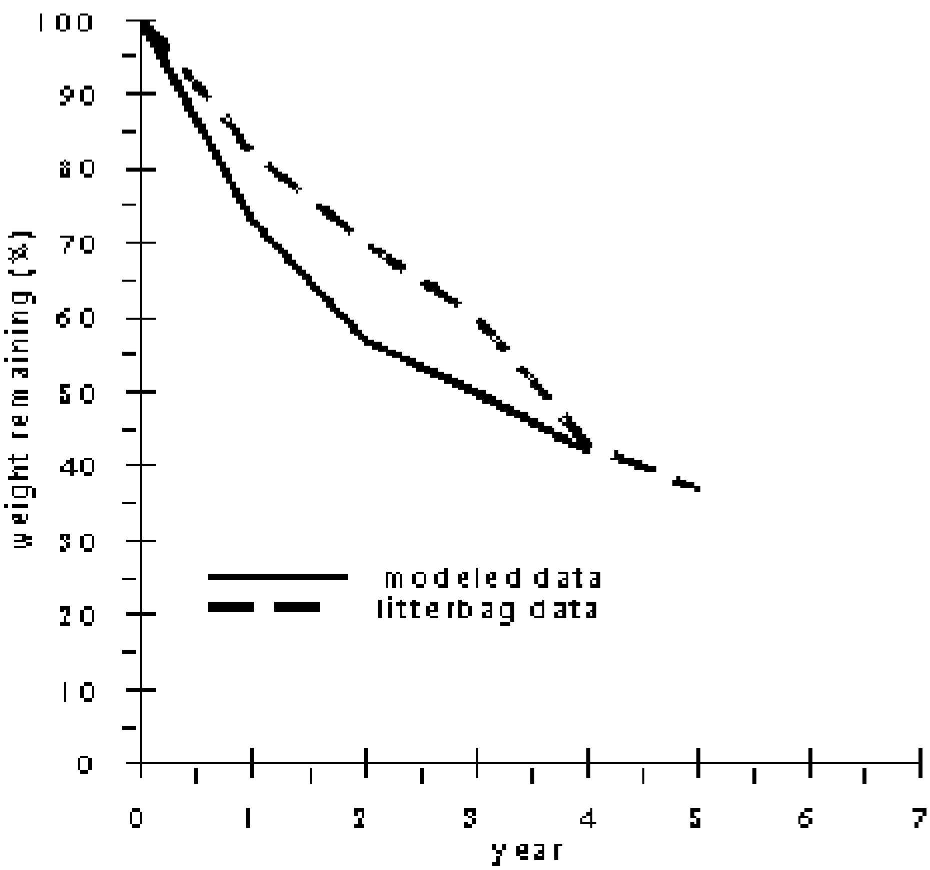


Figure 2. Comparison of foliage decomposition between the modeled cohorts and a set of litterbag data from the floodplain mature white spruce site.

The last set of results that is important to present at this time is the carbon flux, either capture or release, from the individual grid cells and the average for the entire validation plot. This can be presented as a three dimensional map of the carbon capture for the entire validation test site on a square meter basis (Figure 3). The average carbon capture for this stand was 77 grams per square meter for a single year. Total carbon capture represents 155 kg for the entire plot (45 m x 45 m). Carbon capture was found throughout the entire plot but only in the cells that contained a tree. Carbon release was found in cells in which trees were not present. Release was due to decomposition of the forest floor and mineral soil organic matter. For a mature stand this represents one of the best estimates of carbon capture for the boreal forest because of the inclusion of moss in the understory and the inclusion of root growth for the trees present in the model.

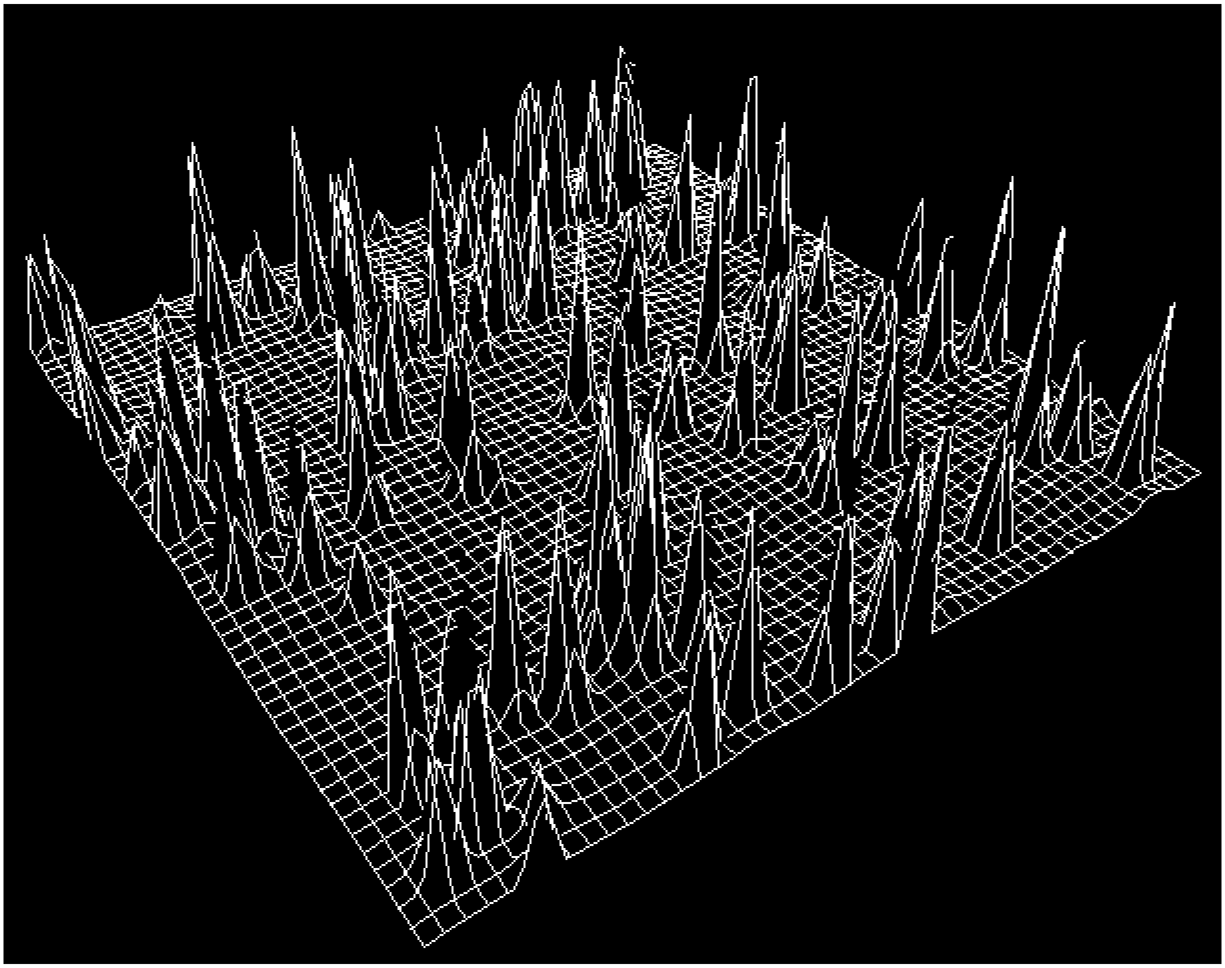


Figure 3. Graphical representation of the carbon captured in a mature white spruce stand on the floodplain of interior Alaska. The highest values reported were 5542 gms/m*m with a minimum value of -288 gms/m*m. The positive value represents carbon capture by trees and the negative value represents carbon release through decomposition of the forest floor and mineral soil organic matter. The grid cell size is one meter.

The estimates for ecosystem carbon uptake were less than those reported by Bonan (1992). He estimated that the trees captured about 1580 gms /m*m in a year and that including moss and microbial respiration the net capture was also about 1580 gms/ m*m per year. This is about 10 times higher than the estimate in the GAFED model. The GAFED model is also estimating carbon release from the forest floor and mineral soil at about 106 gms/ m*m per year (the range is 18 to 230) while Bonan (1992) estimated about 200 gms /m*m per year. The indication from this analysis is that both of the estimates are reasonably correct but neither can be accurately moved to a landscape basis without a large overestimate in one case or an underestimate in the other. The need to move to the landscape with the differences in vegetation types accurately portrayed should be obvious.

The next set of work for the GAFED model will be to verify the carbon dynamics for a number of additional sites within the Taiga LTER program. In addition a verification effort for the landscape level with the data sets that are available for the area surrounding the

Bonanza Creek Experimental Forest will be put together. The final result will be the ability to start to develop accurate estimates of the carbon flux across the forested landscape of interior Alaska.

REFERENCES

Agren, G. I. 1983. Nitrogen productivity of some conifers. Can. J. For. Res. 13, 494-500.

Agren, G. I. 1985. Theory for growth of plants derived from the nitrogen productivity concept. Physiol. Plantarum 64:17-28.

Agren and Bosatta 1987. Theoretical analysis of the long-term dynamics of carbon and nitrogen in soils. Ecology. 68:1181-1189.

Bonan, G. 1992. Physiological controls of the crbon balance of boresl forest ecosystems. Can J. For. Res. 23:1453-1471.

Bosatta, E. and G. I. Agren. 1985. Theoretical analysis of decomposition of heterogeneous substrates. Soil Boil Biochem. 17:601-610.

Gallant, A. L., E. Binnian, J. Ornernik, and M. Shasby. 1995. Ecoregions of Alaska. Geological Survey Professional Paper XXXX. 143 pp.

Hammond, T. and J. Yarie. In Press. Spatial prediction of climatic state factor regions in Alaska. Ecoscience

Ingestad, T. 1977. Nitrogen and plant growth. Maximum efficiency of nitrogen fertilizers. Ambio 6:146-151.

Ingestad, T. 1979a. A definition of optimum nutrient requirements in birch seedlings. III. Influence of pH and temperature of nutrient solutions. Physiol. Plant. 46:31-35

Ingestad, T. 1979b. Nitrogen stress in birch seedlings II. N, K, P, Ca, and Mg nutrition. Physiol Plant. 45:149-157.

Ingestad, T. 1980. Growth, nutrition and nitrogen fixation in grey alder at varied rate of nitrogen addition. Physiol. Plant. 50:353-364.

Ingestad, T. 1981. Nutrition and growth of birch and grey alder seedlings in low conductivity solutions and at varied relative rates of nutrient addition. Physiol Plant. 52:454-466.

Ingestad, T. and M. Kahr. 1985. Nutrition and growth of coniferous seedlings at varied relative nitrogen addition rate. Physiol. Plant. 65:109-116.

Setzer, T. S. 1987. Timber resource statistics for the Porcupine Inventory Unit of Alaska, 1978. USDA Forest Service, res. Bull., PNW-RB-141. 32 pp.

Van Cleve, K., L. Oliver, and R. Schlentner. 1983b. Productivity and nutrient cycling in taiga forest ecosystems. Can.J. For. Res. 13:747-766.

Viereck, L. A., C. T. Dyrness, A. R. Batten, K. J. Wenzlick. 1992. The Alaska vegetation classification. Pacific Northwest Research Station PNW GTR-286.

Yarie, J. 1983. Forest community classification of the Porcupine River drainage, interior Alaska, and its application to forest management. USDA Forest Service, Gen. Tech. Rep. PNW-154. 68pp.

Yarie J. and K. Van Cleve. 1996 (in press). Effects of carbon, fertilizer and drought on foliar nutrient concentrations of taiga tree species in interior Alaska. Ecol. Applic.

Yarie, J. and K. Van Cleve. 1983. Biomass and productivity of white spruce stands in interior Alaska. Can J. For. Res. 13:767-772.

Yarie, J. (in review). Nitrogen productivity of Alaskan tree species at an individual treee and landscape level. Ecology

---

Dr. John Yarie, Associate Professor of Silviculture, Department of Forest Sciences, School of Agriculture and Land Resources Management, University of Alaska Fairbanks. Fairbanks, Alaska 99775. Phone: (907) 474-5650; FAX: (907) 474-6567; email: jyarie@forest.lter.alaska.edu

Eric J. Gustafson and Robert H. Gardner

Dispersal and mortality in a heterogeneous landscape matrix

ABSTRACT: Metapopulation theory predicts that high rates of patch colonization by dispersing individuals exert a stabilizing effect on metapopulation dynamics. It is not understood how the spatial heterogeneity of land cover within the landscape matrix affects the movement of dispersers among patches of preferred habitat, nor how disperser mortality interacts with matrix structure. We simulated the dispersal of organisms from forest patches within heterogeneous agricultural landscapes with a series of self-avoiding random walkers on a digital land cover grid. Each habitat type was assigned *a priori* a probability that the SAW would enter that habitat type. Each individual began the dispersal process on a random site at the edge of a deciduous forest patch and was allowed to move until it reached a different deciduous forest patch. Four mortality functions were simulated: 1) constant probability of mortality at each time step, 2) increasing probability of mortality with each time step, 3) the probability of mortality was related to the habitat type occupied by the SAW at each time step, and 4) no mortality. The results of the sensitivity analysis showed that mortality markedly reduced emigration success even at low mortality rates. Emigration and immigration rates between two patches were not always symmetrical, due to a lack of symmetry in the effect of landscape pattern on disperser movement. Mortality exerts a powerful negative pressure on emigration rates, and the form of the mortality function has a non-trivial effect on simulated dispersal dynamics.

INTRODUCTION

Many landscapes can be characterized as isolated islands of semi-natural habitat within a matrix of diverse land cover types. The equilibrium theory of island biogeography (MacArthur and Wilson, 1967) and the development of metapopulation models (Levins 1970) have allowed the properties of population persistence in space and time to be described as a function of colonization and extinction rates. Studies have shown that patch colonization by dispersing individuals exerts a stabilizing effect on metapopulation dynamics (Kareiva and Anderson 1988), but it is not understood how the spatial arrangement of land cover in the matrix between habitat patches affects the movement of dispersers among patches. It is unreasonable to expect that dispersal routes will be the shortest distance between two habitat islands. Some habitats may be perceived as more hostile to movement than others, and the degree of hostility could be quite important in determining disperser movement patterns. Landscape structure is in fact likely to produce barriers to movement in certain directions, forcing dispersing individuals to concentrate movement within restricted corridors (Hansson 1991) that may not be obvious to human observers.

Mortality of dispersing individuals influences the rate of colonization of habitat patches by eliminating potential colonists. Patches surrounded by land uses where disperser mortality is relatively high would be expected to have a lower colonization rate than those connected to a source of colonists by a travel route consisting of habitats where disperser mortality is relatively low. Although disperser mortality is believed to be related to the type and quality of travel habitat (Wiens et al. 1993), insufficient evidence has been obtained to elucidate any potential interaction between matrix structure and disperser mortality, and their impact on patch colonization rates. If matrix structure and mortality risk do interact to favor the use of those travel routes that minimize mortality risk, then the configuration of dispersal corridors can become an emergent property of the landscape, and simulation modeling may provide insight into the nature of these interactions.

Dispersal in heterogeneous landscapes can be simulated using an individual-based model approach (Kareiva and Shigesada 1983, Gardner et al. 1989), with the results of numerous individuals summarized to estimate colonization rates between patches (Gustafson and Gardner 1996). The simulation approach is flexible, allowing many different movement rules (i.e., step-size of disperser and interaction with landscape), and their interactions with heterogeneous landscapes to be systematically examined. In a previous study, we used an individual-based model of self-avoiding random walkers to estimate the probabilities of transfer of dispersers between habitat patches that were embedded within an agricultural matrix (Gustafson and Gardner 1996). We found that transfer rates between patches were not symmetrical, and that dispersal corridors were diffuse. However, to avoid confounding the effects of matrix heterogeneity, we did not explicitly account for mortality of dispersers.

In this paper, we used this model of self-avoiding random walkers to estimate the probabilities of transfer of dispersers between habitat patches when disperser mortality was explicitly modeled. The model was used to: (1) estimate the effect of different mortality functions on the transfer probabilities between habitat patches, and (2) determine the interaction between disperser mortality and matrix structure.

METHODS

Organisms moving in a heterogeneous landscape were simulated as first-order self-avoiding random walkers (SAW) traversing a gridded surface (Stanley 1986, Bunde and Havlin 1991). Movement was self-avoiding in that the SAW was not allowed to return to a previously visited cell until after >2 steps, which helped prevent the walker from getting trapped in an isolated favorable cell.

The spatial structure of the landscape was represented as a digital land cover map, with each matrix cover type *i* assigned *a priori* a probability, m_i, that the SAW would enter that cover type. We assumed that m_i declined linearly across the 10 habitat types, from the land cover with the most complex vegetation structure to the least hospitable to vegetation (Table 1). Each individual began the dispersal process on a random site at the edge of a deciduous forest habitat patch - the preferred habitat type for this simulated species. Movement across the gridded landscape occurred as a sequence of steps to adjacent cells. At each step, the value of m_i for the previously visited site was temporarily set to zero to prevent back-tracking (i.e., a self-avoiding walk). A cumulative probability distribution of the values of *m* associated with the remaining 7 cells was constructed, such that the probability of moving into cell *i* (p_i) was a function of the land cover types in the 7 cells (*j*):

A uniform pseudo-random number was then selected to choose the cell to which the SAW moved. The map edges were assumed to be reflective, (i.e., the probability of entering sites beyond the map boundaries was 0.0). Each walker was allowed to move until it reached a deciduous forest patch different from the one it left initially, or died. An upper limit (s_{max}) of 5000 steps was set for each SAW, in which case it was assumed to have reached its exploration limit and to have failed to disperse. For each deciduous forest patch on a landscape, 10,000 SAWs were simulated and the origin and destination patch of each individual was tabulated.

Table 1. Movement (m_i) and mortality ($p_{mort(i)}$) probabilities for each habitat type *i*, that were used to simulate disperser movement and mortality.

	Probabilities	
Habitat (*i*)	Movement, m_i	Habitat-dependent Mortality, $p_{mort(i)}$
Water	0.001	0.05
Road, developed	0.01	0.005
Bare soil, dry	0.02	0.002
Bare soil	0.1425	0.00033
Bare soil, moist	0.265	0.00018
Young row crop	0.3875	0.00012
Short, grassy	0.51	0.00009
Medium, grassy	0.6325	0.00007
Tall, grassy	0.755	0.00006
Conifers	0.8775	0.00005
Deciduous forest	1.0	N/A

Four disperser mortality scenarios, exhibiting a gradient of decreasing exposure to mortality risk, were simulated: 1) constant probability of mortality, 2) time-dependent increase in mortality over simulated time, 3) habitat-dependent probability of mortality, and 4) no mortality. For the constant mortality case, the probability of mortality (p_{mort}) was held constant throughout the movement of each SAW. In the time-dependent mortality case, p_{mort} was initially very low, but increased linearly at each time step. This is analogous to the situation where a disperser becomes more at risk of mortality as its energy and water reserves are depleted. In the habitat-dependent case, p_{mort} was related to the habitat type of the cell where the SAW was located at each step. We assumed that mortality by predation was less in more complex habitats. The mortality risk associated with each of the habitat types was determined by assuming that mortality risk in each habitat *i* was inversely proportional to m_i. Thus, SAWs were more likely to enter habitats where mortality was relatively low. To allow meaningful comparisons, we attempted to make the average p_{mort} equal across the 3 mortality scenarios. $p_{mort} = 0.005$ for the constant mortality case, and p_{mort} for the time-dependent mortality case had a mean of 0.005 when the full 5000 steps were taken. The average p_{mort} for a SAW under the habitat-dependent case could not be calculated *a priori*, since the SAW will preferentially enter habitats with lower p_{mort} values, but the average of the p_{mort} values in Table 1 equal 0.005. At the end of each step, if a uniform random deviate was <= $p_{mort,}$ the SAW was terminated, and the number of steps taken and the habitat last visited was recorded.

Three agricultural landscapes in northwestern Indiana of 36 km2 were selected for the simulations from a larger classified Thematic Mapper (TM) image (pixel width = 30 m) collected on 7 June 1989. The image was classified to produce a land cover map as described in Gustafson and Parker (1994), to produce the classes delineated in Table 1. TM land cover data were selected to provide realistic landscape patterns.

These landscapes provide examples of the patchy distribution of deciduous forest habitat in agricultural landscapes, each with noticeably different spatial structures. The first sample (MINE) contained a gravel-mining operation, with associated roads, lagoons and bare parent material visible in the center of the study area (Figure 1). The second sample (RIVER) had a large river running through it (not shown), and the third (AGRI) was in a rural agricultural area with no major roads (not shown). The configuration of deciduous forest patches was quantified by calculating the area, perimeter, and isolation (distance to the nearest other deciduous forest patch) of each patch.

Figure 1

Figure 1. Land cover map of the MINE study area and the ID numbers for each deciduous forest patch on the study area.

To determine the sensitivity of the model to changes in model parameters, we exercised the model on the RIVER landscape by systematically varying $p_{mort,}$ $s_{max,}$ and the mortality factor. Simulation results were summarized as a matrix, **A**, of dispersal success from each patch, *j*, to every other patch, *i*. The proportion of walkers to successfully emigrate from patch *j* to any other of the *n* deciduous forest patches on the landscape was calculated as E_j = SUM A_{ij} , i = 1...*n; i* < > *j*. A landscape dispersal coefficient was calculated as *LDC* = [SUM ($E_j \times S_j$)] / SUM S_j , j = 1...*n*, where S_j is the area of patch *j*. *LDC* summarizes the success rate of dispersing individuals across an entire landscape. Landscapes with many impediments to dispersal would have low values of *LDC,* while landscapes that facilitate dispersal would have high values of *LDC*.

To assess the relative impact of various characteristics of landscape structure on dispersal success, we related patch size, perimeter, isolation, and the mortality related parameters to the probability of successful emigration (E_j) using ANOVA. Visualization of the movement patterns across the landscape was achieved by tabulating the frequency of visitation of successful dispersers to each grid cell on the map and displaying the visitation frequency as a continuous color scheme. Inspection of this map revealed corridors and barriers on the landscape, defined as areas where movement was more or less frequent, respectively.

RESULTS

The results of the sensitivity analysis showed that constant mortality markedly reduced the LDC even at low values of p_{mort} (Figure 2). The habitat-dependent case allowed dispersers to avoid mortality by preferentially choosing habitats with lower mortality risk, resulting in a more linear decline in *LDC* as the mortality rate increased. The time-dependent case was intermediate, since mortality was lower than the average p_{mort} at the beginning of each random walk, allowing more dispersers to succeed while mortality rates were low. Increasing s_{max} served to increase *LDC* by allowing dispersers more opportunity to successfully disperse, although as p_{mort} increased, dispersers tended to die before taking s_{max} steps.

Figure 2. Results of sensitivity analysis showing the relationship between the Landscape Dispersal Coefficient (*LDC)*, and the maximum number of steps the SAW was allowed to take (s_{max}) and the average mortality rate (p_{mort}), for each mortality scenario.

Emigration and immigration rates between two patches were not always symmetrical, that is, A_{ij} does not necessarily equal A_{ji} . This effect is due, in large part, to a lack of symmetry in landscape pattern. For instance, an isolated patch may have a high proportion of dispersers reach its nearest neighbor, but if its neighbor has a number of closer neighbors, only a small proportion of the neighbor's dispersers will reach the more isolated patch. The structure of the landscape matrix may also cause asymmetry in transfer probabilities. Barriers to dispersal may act as a funnel favoring dispersal in one direction, while deflecting movement in the opposite direction. Differences in mortality functions did not change the relative magnitude of asymmetries between patches, only the absolute values of the transfer rates.

Maps that show the relative frequency of visitation of each cell by successful dispersers enable visualization of the most likely route that a disperser will travel between patches (Figures 3a-3c). Notice that these dispersal routes (corridors) tend to be diffuse. Mortality serves to eliminate those dispersers that choose indirect routes under the constant and time-dependent mortality scenarios, and favors those that use habitats with low mortality risk under the habitat-dependent mortality scenario (Figure 3).

Figure 3a - MINE
Figure 3b - RIVER
Figure 3c - AGRI

Figure 3. Maps of each landscape showing the relative frequency of visitation of each cell by successful dispersers under each mortality function.

The habitat-dependent mortality case always had a relatively high emigration success because dispersers favored habitats with the lowest p_{mort}. The time-dependent and constant mortality cases resulted in declines in emigration success (E_j) from a given patch, a consequence of reduced opportunities for dispersers to avoid mortality (Figure 4). However, the magnitude of the decline was dependent on the configuration of the neighboring patches, and the structure of the matrix. For patches with a favorable route (i.e., corridor of habitat with low p_{mort}), or a short route to another patch, the decline in E_j was less than for patches with distant neighbors or unfavorable routes. As an example of this, compare patches 9 and 12 on the RIVER landscape (Figure 4). They have similar values of E_j under no mortality, but patch 9 has lower E_j values under the time-dependent and constant mortality cases because it is more isolated, and has relatively unfavorable routes to its neighbors. Exposure to mortality risk by an individual SAW was a function both of the mortality function and the configuration of the habitat.

Figure 4. Effect of changing the mortality function on emigration success from deciduous forest patches on the three study areas. The MINE study area is illustrated in Figure 1.

The ability of dispersers to move freely from patch to patch on a given landscape (as measured by *LDC*) was reduced under the time-dependent and constant mortality scenarios (Figure 5). Furthermore, the differences in *LDC* among landscapes with different patch and matrix configurations became less under these scenarios, such that *LDC* is essentially the same for each landscape under the constant mortality scenario (Figure 5).

Figure 5. Overall dispersal success as measured by the Landscape Dispersal Coefficient (*LDC*) on each landscape under the four mortality scenarios.

The analysis of variance showed that the greatest effect on E_j was due to differences in the mortality functions simulated, explaining nearly 50% of the total variance in the total sums of squares (Table 2). The distance to the nearest neighbor patch explained an additional 25% of the variance. Also significant was the inverse relationship between the average number of steps taken by successful dispersers and E_j . In the face of mortality, the more steps a disperser must take to reach another patch, the less likely that it will survive to successfully disperse.

Table 2. Analysis of the variance in emigration success (E_j) rates attributed to landscape structure and mortality functions on the three study landscapes.

Source	d.f.	SS	F	Prob > F	R^2
Patch size	1	0.3532	23.03	0.0001	
Patch perimeter	1	0.0799	5.21	0.0242	
Patch isolation	1	2.8576	186.31	0.0001	
Mortality function	1	5.6052	365.44	0.0001	
Steps / mortality [1]	1	0.0080	0.52	0.4725	
Steps / disperse [2]	1	0.7116	46.39	0.0001	
Error	117	1.7946			
Total	123	11.4101			0.84

[1] Average number of steps taken by SAWs that suffered mortality.
[2] Average number of steps taken by successful dispersers.

DISCUSSION

The probability of population extinction within a single isolated habitat patch increases as an inverse function of the size and quality of the patch (Kareiva 1990, Hanski and Gyllenberg 1993). Therefore, the long-term persistence and stability of populations within fragmented landscapes depends on the rate of dispersal and recolonization among habitat patches (i.e., the "rescue effect" of Brown and Kodric- Brown 1977). Unfortunately, the process of dispersal is poorly understood and difficult to measure (Gaines and Bertness 1993), with the result that metapopulation studies have generally ignored the rich interaction of species with the spatial patterns of the landscape matrix that separates habitat patches (Wiens et al. 1993).

The results of our study show that mortality exerts a powerful negative pressure on emigration rates, and that the form of the mortality function has a non-trivial effect on simulated dispersal dynamics. We found that 50% of the variability in dispersal success can be accounted for by differences in mortality functions (Table 2). This suggests that detailed investigation into the nature of disperser mortality is warranted in studies of metapopulation dynamics. An assumption of a constant risk of disperser mortality in landscapes with a heterogeneous matrix may bias the results.

The addition of mortality in these simulations did not reduce the prevalence of asymmetrical transfer rates between patches that we found in our previous study (Gustafson and Gardner 1996). This supports the idea that such asymmetries are caused by irregularities in the arrangement of habitat patches and the structure of the matrix, which causes transfer rates from a patch to be different than transfers to it. However, mortality tends to depress all transfer rates, and may have a significant effect on patches with marginal immigration rates.

The rates produced by our individual-based dispersal model also provide a means for improving estimates of patch-to-patch transfers for metapopulation and source-sink models. Metapopulation and source-sink models generate predictions about within-patch population dynamics and the relationship of discrete populations by dispersal, but usually assume that dispersal is unaffected by the intervening landscape matrix. Our cellular automata model, on the other hand, explicitly estimates the probability that a disperser from a given patch will successfully arrive at other patches on the landscape, but does not consider within-patch dynamics or estimate the actual number of dispersing individuals. The marriage of these approaches appears promising, and would be a means to more explicitly represent spatial dynamics within metapopulation models. The linkage of spatially-referenced land cover data with an individual-based model of dispersal holds promise for landscape ecological research as a means to link ecological pattern and ecological process.

ACKNOWLEDGMENTS

We thank S.M. Tang for critical comments on earlier drafts of the manuscript. This research was funded in part by the North Central Forest Experiment Station, Appalachian Environmental Laboratory - Center for Environmental and Estuarine Studies, and an EOSAT Award for Applications of Landsat Thematic Mapper Data.

REFERENCES

Brown, J. H., and Kodric-Brown, A. (1977) Turnover rates in insular biogeography. Ecology 58:445-449.

Bunde, A., and Havlin, S. (1991) Percolation I. Pages 51-96 *in* A. Bunde, and S. Havlin, editors. Fractals and Disordered Systems. Berlin: Springer-Verlag.

Gaines, S. D., and Bertness, M. (1993) The dynamics of juvenile dispersal: why field ecologists must integrate. Ecology 74: 2430-2435.

Gardner, R. H., O'Neill, R. V., Turner, M. G., and Dale, V. H. (1989) Quantifying scale-dependent effects of animal movement with simple percolation models. Landscape Ecology 3:217-227.

Gustafson, E.J., and Parker, G.R. (1994) Using an index of habitat patch proximity for landscape design. Landscape and Urban Planning 29:117-130.

Gustafson, E.J., and Gardner, R.H. (1996) The effect of landscape heterogeneity on the probability of patch colonization. Ecology 77(1):xx-xx. In press.

Hanski, I., and Gyllenberg, M. (1993) Two general metapopulation models and the core-satellite species hypothesis. American Naturalist 142:17-41.

Hansson, L. (1991) Dispersal and connectivity in metapopulations. Biological Journal of the Linnean Society 42:89-103.

Kareiva, P. (1990) Population dynamics in spatially complex environments: theory and data. Philosophical Transactions of the Royal Society of London B 330:175-190.

Kareiva, P., and Anderson, M. (1988) Spatial aspects of species interaction: the wedding of models and experiments. Pages 38-54 *in* A. Hastings, editor. Community ecology. New York: Springer-Verlag.

Kareiva, P., and Shigesada, N. (1983) Analyzing insect movement as a correlated random walk. Oecologia 56:234-238.

Levins, R. (1970) Extinction. Pages 77-107 *in* M. Gerstenhaber, editor. Some mathematical problems in biology. Providence, Rhode Island: American Mathematical Society.

MacArthur, R.H., and Wilson, E.O. (1967). The theory of island biogeography. Princeton, New Jersey: Princeton University Press.

Stanley, H. E. (1986) Form: an introduction to self-similarity and fractal behavior. Pages 5-16 *in* H. E. Stanley,. and N. Ostrowski, editors. On Growth and Form. Dordrecht, Netherlands: Martinus Nijhoff.

Wiens, J. A., Stenseth, N. C., Van Horne, B., and Ims, R. A. (1993) Ecological mechanisms and landscape ecology. Oikos 66: 369-380.

Eric J. Gustafson
USDA Forest Service, North Central Forest Experiment Station
Forestry Sciences Laboratory
Rhinelander, WI 54501
Telephone: (715) 362-1152
FAX: (715) 362-1166
E-mail: ericgus@newnorth.net

Robert H. Gardner

Appalachian Environmental Laboratory
Center for Environmental and Estuarine Studies
University of Maryland System
Frostburg, MD 21532
Telephone: (301) 689-3115
FAX: (301) 689-8518
E-mail: gar@ael.umd.edu
Page 9

Integrating ecological tools with remotely sensed data: modeling animal dispersal on complex landscapes

Gillian Bowser--National Park Service, Denver, Colorado and the University of Missouri-St. Louis

---

INTRODUCTION

Using geographic information systems (GIS) to explore the spatial relationships of animal populations is a relatively new field for ecologists (Johnson 1990, Scott et al. 1993) and one untouched by population geneticists. GIS, as an environmental modelling tool, evolved from simple beginnings as a mapping program to a modelling and analysis engine for a variety of different disciplines (Goodchild 1993). GIS is well-established in habitat-based studies of animal populations to analyze remotely-sensed databases (Johnson and Naiman 1990) and as a predictive tool for animal or plant species distributions (Scott et al. 1993, Jensen et al. 1992). In addition, GIS is now used to create databases, manipulate spatially-explicit surfaces to represent specific parameters, and to displace spatial relationships through simulation modelling, hydrologic constructs, and species relationships (Keller 1990; Aspinall and Veitch 1993). One application still unexplored with GIS, despite the importance of spatial heterogeneity, is the animal population dynamics as expressed by genetic parameters.

Spatial heterogeneity is a complex phenomena involving patterns or mosaics of habitats at a landscape level (Kolasa and Rollo 1991, Turner and Gardner 1991, Dunning et al. 1992). The causes and consequences of spatial heterogeneity can be measured by examining patterns of variation in the population dynamics of an organism as recorded in the distribution and exchange of individuals. Environmental variables, as independent variables either abiotic or biotic, can control variation within "dependent taxonomic groups" (Karieva 1990; Green 1994) by influencing the movement of individuals among population units. Thus, the influence of those environmental features should be reflected in population measures such as genetic differentiation (Wayne et al. 1990; Patton and Smith 1992; Opdam 1991; Mitchell-Olds 1992; Lande 1988; Stewart et al 1993).

Two features of population dynamics, gene flow and genetic drift, are particularly sensitive to the spatial milieu in which populations are embedded. Populations persist in part through the exchange of individuals (gene flow) and maintance of genetic fitness (Slatkin 1981, 1985; Leberg 1990). In contrast, as populations become isolated and genetic mistakes occur and accumulate, the population drifts towards genetic differentiation and speciation (Leberg 1990, Slatkin 1993). As gene flow implies the migration of individuals between populations, genetic drift is a measure of the isolation or separation among populations (Wright 1965, Slatkin 1985, Weir and Cockerham 1984). In island models of genetic differentiation, the genetic separation among populations is analogous with their physical distances (Slatkin 1993). Genetic distance, as a measure of accumulated differences in allele frequencies among populations, is a single number (Jacquard 1970, Weir 1990) that characterizes populations by the genotype frequencies among members, condensed from information on the frequencies of various genes or alleles for the entire population (Jacquard 1970). Genetic distance, as a representation of linear distance, is an estimate as to how populations spatially organize themselves and is a data reducing device (Weir 1990) that can infer population groupings--phylogenetic trees or similarity cluster diagrams (Rogers 1991)--comparable to groups created by physical distances. Thus, euclidean distance (the linear distance or difference between two points characterized by many variables) approximates genetic distance when calculated across a simple landscape (Slatkin 1993).

In this study, theorectical dispersal paths were calculated using GIS first to build landscapes based on individual environmental features, then to redefine these landscapes with pre-conceived concepts on the ecological perspectives of a species. Spatially-explicit analyses use the physical path a dispersing animal follows across these newly-created landscapes to explain patterns of genetic differentiation among populations. The hypotheses for this study address the paths drawn across these ecologically-created landscapes:

- Environmental features (surfaces based on one type of environmental feature--roads, streams, elevation, etc.) impact animal dispersal paths. Paths drawn based on this impact will better reflect the genetic distances among populations than the Euclidean distances.
- Environmental surfaces act synergistically. Individual surfaces do not explain genetic distances as well as combinations of multiple surfaces.

Protein electrophoresis of alloyzmes from a fossorial colonial rodent (*Cynomys ludovicianus*) was used to determine genetic distances among different populations in Badlands National Park (BADL), South Dakota. Genetic distance was used as response variable in regression models where the predictors were dispersal paths created by Geographic Resources Analysis Support System (GRASS) version 4.1 GIS software across a variety of different environmental surfaces. Dispersal paths were tested in two steps: first each GIS-created path was compared with the Euclidean distance using two-way analysis of variance (ANOVA); second, each path was tested against the genetic distance (corrected to match the scale of the distance in meters) using multiple linear regression to find which set of paths best explained genetic distances.

METHODS

The study organism

Black-tailed prairie dog (*Cynomys ludovicianus*), a fossorial colonial rodent, populations from BADL were chosen because their colonies leave distinct scars on the landscape, easily mapped from low-level aerial photography (Schenbeck and Myhre 1990). These colonies can be aged by tracking scars back through time, using historical aerial photography, to minimize possible extinction and colonization effects on the genetic differentiation of focus colonies.

Fourteen colonies were identified on aerial photography; seven were chosen for this study based on their age, stability, and size.

Genetic data

Genetic data was gathered from each colonies by live-trapping or destroying 20 individuals, each from separate family groups or coteries (Chesser 1983, Daley 1992). Blood samples were collected in hemphanized 250 microliter tubes and immediately stored on dry ice to prevent protein decay. Samples were shipped to the University of Missouri-St. Louis and analyzed for 21 presumptive proteins using horizontal starch-gel protein electrophoresis on three different (pH 7.1, 8.0, 9.1) buffer systems (Richardson et al. 1986, Pasteur et al. 1988, Aquaah 1990). Allele frequencies from seven polymorphic proteins were calculated along with genetic distances based on Modified Rogers' algorthim (Rogers 1991) using BIOSYS-1 software (Swofford and Selander 1981) on an IBM PC.

The Model

Models were created to build spatially-explicit dispersal paths among populations on complex ecologically-defined landscapes. The models themselves are translations of digital spatial data incorporated into GRASS and transformed by mathematical and ecological parameters using UNIX Bourne-shell scripts. Each model consists of several elements: (1) a transforming surface--a GIS layer based on a single environmental feature such as elevation, streams, roads, etc.; (2) a weighting protocol--weights are assigned to each element on the transforming surface based on a preconceived concept on the ecological impact of that feature on animal dispersal; (3) a cost surface--a surface created by the GIS that defines each pixel on the transforming surface based on it's cummulative weight (cost) from a chosen starting point like a population center, to a specific endpoint; and (4) a dispersal path--a route drawn across the cost surface based on choosing the lowest cummulative cost to move between two specified points on that surface.

Transforming surface

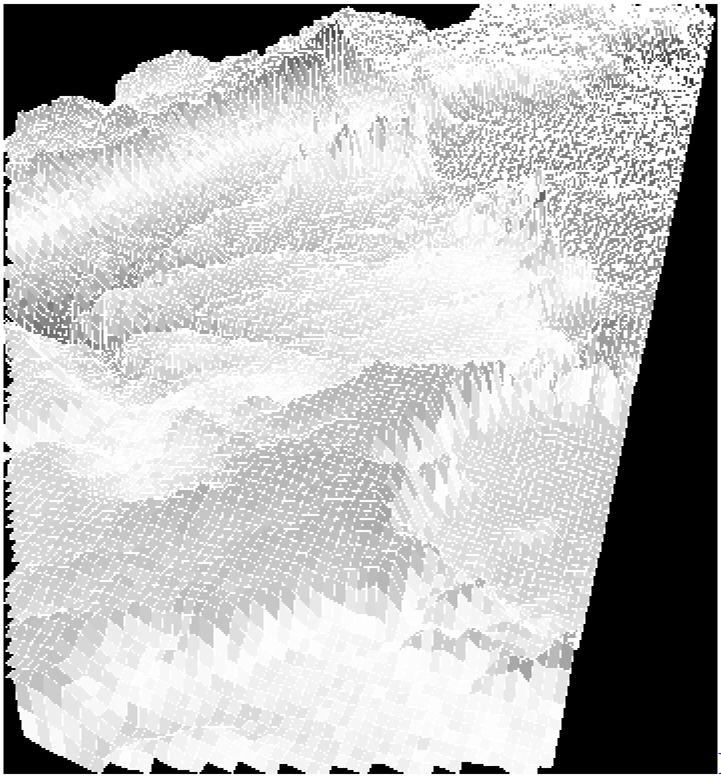


Figure 1 (click here for high res.)

Digital databases used in the models included digital elevation model (DEM), digital line graph (DLG), vegetation (based on Soil Conservation Service DLG data), and landuse boundaries--based on Defense Mapping Agency data. From these databases, slope, elevation, aspect, hydrology, and roads, were extracted and separated into geographically referenced surfaces (Figure 1 elevation surface). These surfaces, such as elevation, represent the biotic environment used as 'seed' surfaces for the dispersal models. Scanned and rectified aerial photography was overlaid on these digital databases and population centers were calculated by screen-digitizing colony scars visable in the photography. Residual Mean Square error (RMS) was calculated for each photograph and recorded along with the RMS of the base photograph or mapand the number of control points used.

Weighing protocol

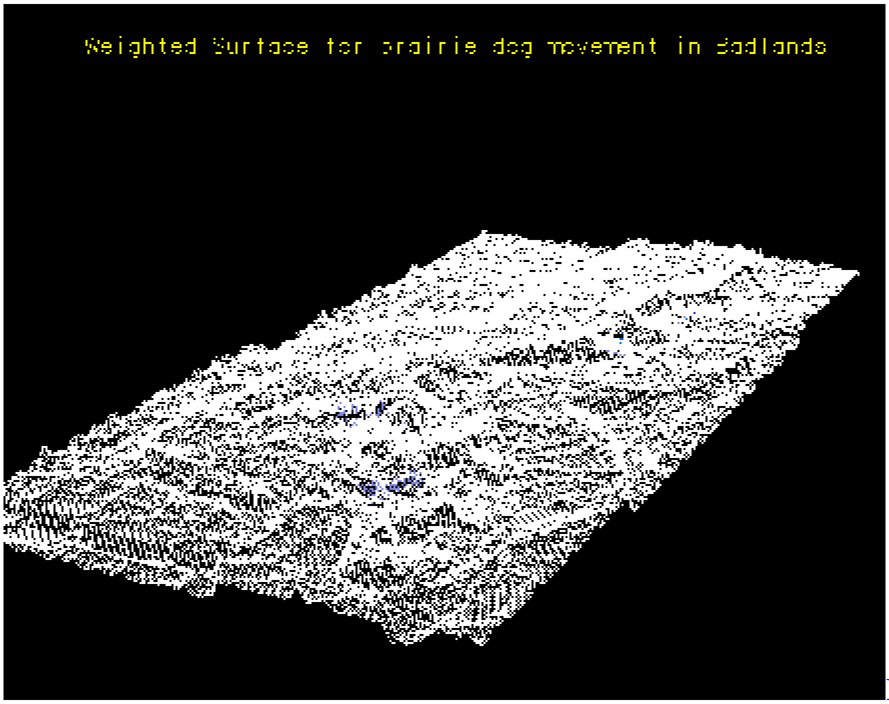


Figure 2 (click here for high res.)

Every feature on a transforming surface was assigned a weight on a geometrically increasing scale, where the weights were based on a pre-conceived concept as to whether that feature consistutes a barrier or deterient to dispersing animals. The new weight surface bears little resemblance to the initial transforming surface as weights are now used to express 'height'. For example paved roads would be assigned a weight higher than a four-wheel drive track, but lower than an interstate highway--thus an interstate appears higher than a paved road on a three-diminsional representation of a transformed surface (Figure 2 weight surface)--giving a different appearance than the original transforming surface (Figure of elevation surface). The weights themselves were arbitrary (chosen in increments of tens or hundreds), and the rate of increase was consistent for all classes on a particular transforming surface. Thus an intermittent stream may have a weight of 10 on the stream transforming surface, and a foot-trail can also have a weight of 10 on the road surface, thus allowing for metric comparisions across surfaces and variables during the model analysis. Any features with no known impact on the ability of an animal to disperse was assigned a zero value. 'No data' values were assigned to missing variables or to areas outside the study area to reduce computer memory requirements. Finally, for surfaces containing habitat parameters (slope, vegetation, aspect) with information on the preferred habitats (based on back-tracking population position through 50 years of aerial photography), all preferred elements were assigned low weights based inversely on their occurrence within population boundaries over time. Thus areas not preferred would be assigned higher weights, but not high enough to completely deter dispersal (a discussion of using weights to compare variables and routes in GIS analysis is in Waggoner 1989; Dr. R. Root, Technology Transfer Center, National Biological Service, personal communication).

Cost Surface

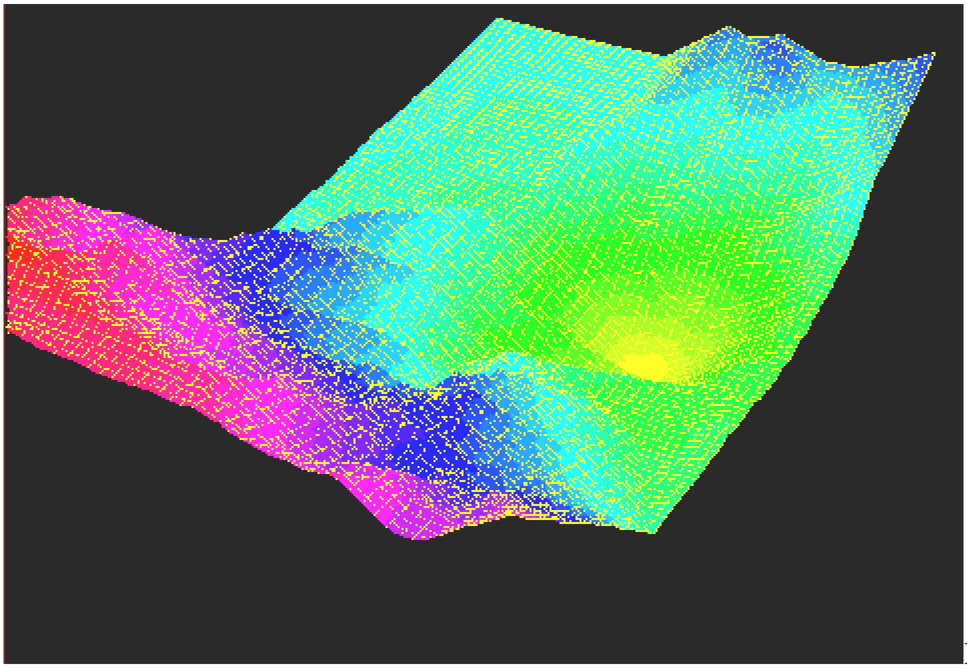


Figure 3 (click here for high res.)

Cost surfaces are representative surfaces of the cummulative cost (based on assigned pixel weights) of moving to each cell on a surface from a defined point (geographic coordinates). Each cell on the surface is reassigned a category value representative of the cost of traversing that cell (Awaida and Westervelt 1993). By using the centroids of each prairie dog colony as a starting point for a cost surface, seven new cost surfaces are generated for each transforming surface (Figure 3 cost surface. The pit in this surface represents the area of lowest cost from the starting point, accumulating cost as the distance from the starting point increases and the program encounters weights on the surface. The cost surface is defined from a starting point, and is analogous to an inverted peak which water will then be drained off of. An extra layer was then added to each surface to represent time. This layer consisted of concentric rings of increasing weights (the size was dependent on the resolution of the cost surface) originating at the starting centroid. These rings simulate time and prevent the program from backtracking or making wide detours. The time surface was meant to incorporate into the dispersal model the increasing risk to animals the longer they are outside colony boundaries.

Dispersal paths

Dispersal paths were created by defining an animals movement as a cummulative sum of cell values crossed as the animal moves between two points. Dispersal paths can be thought of as the path water follows to get to the lowest point on a surface. As water moves down from one point to another, it seeks the path with the least resistence and always chooses the downhill path. A dispersal path for animals is similar, assuming the animal always follows the easiest and quickest route, or 'least-cost' path for dispersal.

There are some construction restraints based on how GIS builds dispersal paths. Paths are constructed using the following equation:

d = (Xwij+Ywij+...)

where the first summation represents moving between two designated points (a and b) on the landscape; *ij* indicates scale of pixels containing features at scale-*i* and resolution-*j*. W is a measure of the weights for feature X or Y and the summuation of those features by scale, resolution, and accuracy will generate **d**--an actual distance going from a to b. Scale and Resolution (ij) are limited by the underlying databases and create a source of error around the placement of an individual point (a or b) for features (X, Y) on a given surface. However, there is also error associated with the calculation of each surface that involves the ability of the GIS to correct the surface and incorporate the data into a coordinate system (based on a reference surface) each with an associated error. For clarity, all of those sources of model construct error are summed into one term--accuracy (Z) for a map layer where the actual path constructed by the model may be represented by:

dr=d/Zx|Px

where the realized path (**dr**) is a function of the constructed path, d, the accuracy of that path for surface x (Zx) given the precision of surface X (P) to an actual ground location. Precision is the degree of detail in reporting a measurement or the arthematic calculation of scaling (Goodchild 1993). In model validating, the terms Z and P will be incorporated into the error term as representative of a sphere of error around each point on the surface that has a diameter determined by Z and P.

Statistical Analysis

All statistical analysis was performed on a UNIX workstation with Splus3.2 software. There were three separate hypotheses tested by the statistical analysis:

- Dispersal paths generated by GIS models do not differ from the euclidean distances among populations.
- Dispersal paths across individual environmental surfaces can predict genetic distances; Single regression models are less effective than multiple models using different combinations of surfaces
- There is significant redundancy among paths drawn by different layers. Independent linear vectors are better able to explain genetic patterns than individual layers.

RESULTS

Colony Placement

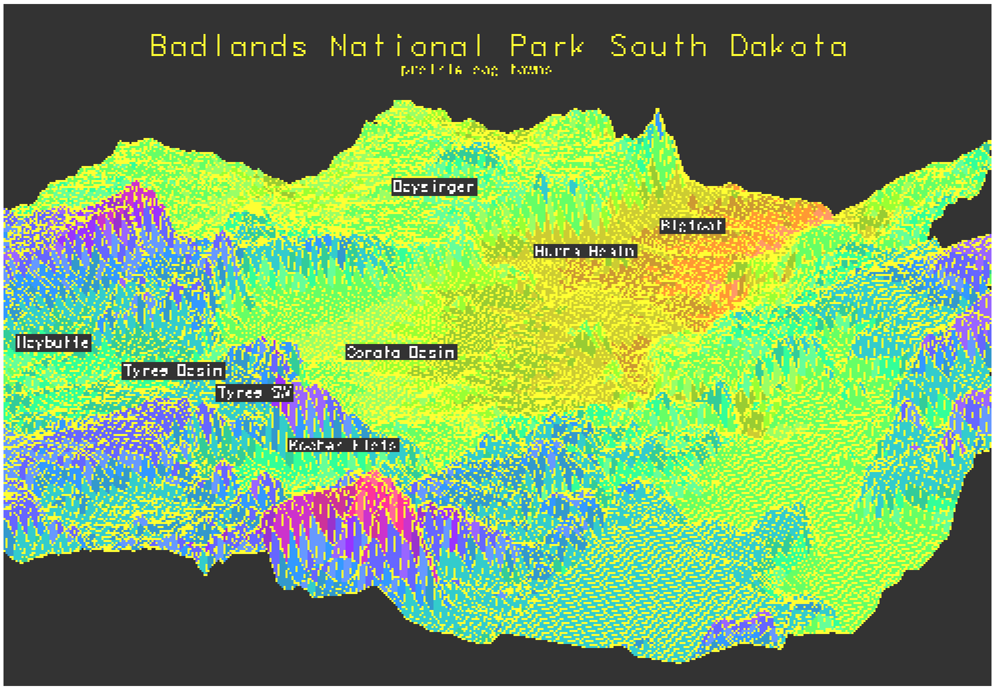


Figure 4 (click here for high res.)

Seven colonies in or adjacent to BADL were used with 42 paths constructed (all possible pairs) per model run (Figure 4 badlands national park . Colony size ranged from 25-70,000 hectares with an estimated mean prairie dog density of 43 per hectare. The seven colonies were all at least 43 years old; some showed signs of budding and expansion (as sources for new populations) while others had neighbors go extinct (sink relationships--Green 1994). Both of these phenomena have implications important to the calculation of genetic distances that were not directly tested by this model.

Photography

Photography was scanned for one meter horizontal accuracy; however, the mean residual square error (RMS) from the photography (1:24,000) to the 1982 base photograph (1:250,000) used for rectification, ranged from 3.5 to 7 meters; while the RMS for the 1982 base to the USGS 7.5 minute topographic quadrant maps was 4.5 meters. As the topographic map had a minimum RMS of 11 meters, the minimum horizontal accuracy was approximately a 15 meter sphere around the centroid locations used to generate paths. This accuracy was considered a source of error in the models analogous to variance around each path constructed by the GIS model.

Single Surfaces versus Euclidean distance

Models constructed from a single transforming surface (single surface models) were significantly different from those constructed using euclidean distance (two-way ANOVA, p=0.05). Some surfaces were quite different from euclidean distance (t-test, p=0.05) with roads (mean path length= 14903 pixels) and streams (mean path length= 15898 pixels) having dispersal paths significantly longer than those created by euclidean distances (mean path length = 13961 pixels). Pixel size for these layers was later adjusted to 30m per pixel to compare with ground measurements.

Single Surface Predictions

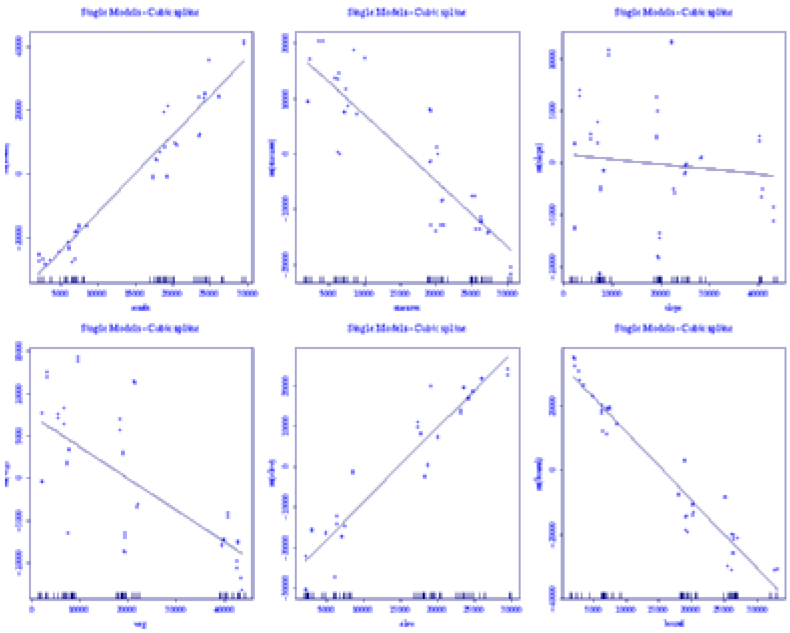


Figure 5 (click here for high res.)

Single surface models did not individually explain genetic distance as a linear response. Each single surface model was tested individually against genetic distance and only two surfaces, roads and boundary were significant predictors of genetic distance (p=0.01, Figure 5 single surface regressions), but the regression models were both poor in their ability to explain the variance around the regression line (multiple r2=0.14). Residual plots on the other surfaces indicated non-linear responses and these models were tested by first applying a spline correction and loess smoother to the data and running loglinear regression. These smoothed models were better predictors (r2=0.33, Figure 5 single surface smoothed regression) suggesting non-linear relationships.

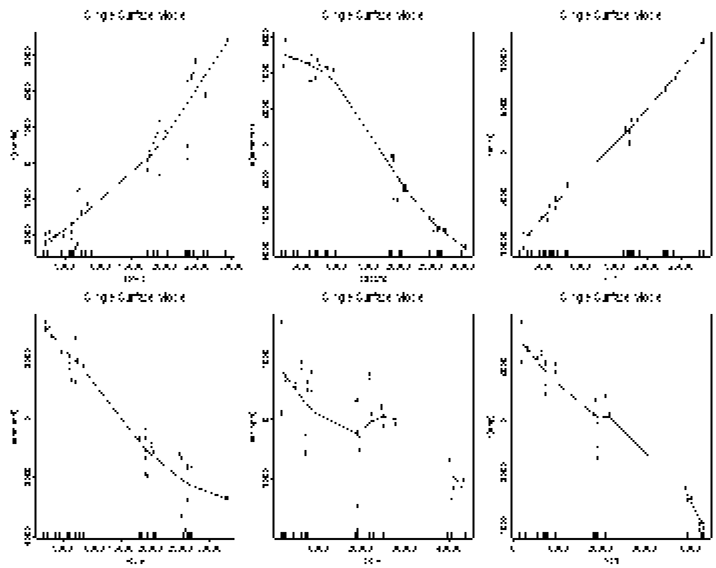


Figure 6 (click here for high res.)

Principal Component Analysis

Principal Component Analysis on five single surfaces showed 65% of the variance explained by the first principal component and vegetation contributed the most to that axis. High multicollinearity in the last two principal components suggests that some of the surfaces showed signs of auto-correlation probably based on use of common centroids for dispersal path construction. Because of this multi-collinearity a multiple regression using a stepwise selection procedure (using Mallows statistic for a selection criteria) was run on the least collinear surfaces and not the complete dataset. From this the best multiple surface run consists of roads, and streams; however the ability to explain variance declines to r2=0.14.

Groupings generated by PCA were weak and did not resemble those obtained by genetic distance alone, nor that expected by euclidean distance. The main source of disagreement appeared to be when GIS calculates paths from one town to another, the reverse path is not always the same length. These path reversals differed in the PCA grouping causing noticeably more scatter than in either the genetic distance or geographic distance groupings. PCA groupings on dispersal paths did have better grouping by starting locations such that loose groups for four of the seven locations were noted.

CONCLUSION

The dispersal paths tested provide some interesting insights to possible environmental influences population genetic characteristics. If population stability for 43 years is sufficient to stabilize genetic relationships (approximately 21 generations), then stable populations should be reflecting the processes of genetic drift and gene flow. For prairie dogs in BADL, gene flow seems to be most influenced by the presence of streams and roads and least influenced by the highly erosive cliffs characteristic of the badlands environment (Figure 1 elevation surface). These results lead to some interesting speculation. The model draws paths of different lengths (for a single pair of locations) depending on the direction of travel; however, if the dispersal paths are examined in light of environmental features, the difference appears to be related to when the drain program encounters a significant barrier. Because a time surface was imposed on top of the landscape, significant barriers encountered at the end of a dispersal path are less likely to divert the animal than one close in (when the time surface has the strongest effect). The ecological significance of this unexpected model artifact is that the placement of some features on a landscape will have differing effects dependent on the distance from the source population when that feature is encountered. To test this hypothesis is outside the current range of these dispersal models but warrants further exploration.

Sources of Error

There are three sources of error in the dispersal models: (1) error associated with the ecological restraints of the animal; (2) error associated with database accuracy and precision; and (3) error due to the calculation of the predictor variable itself.

Ecological Error-- The ecological restraints of the study animal is a source of error and conflict between GIS databases--remotely sensed at large pixel sizes, and a mid-size rodent with a home-range often as small as the minimum pixel size. Prairie dogs weigh an average of 2-5 pounds and rarely disperse distances greater than 10 kilometers (Knowles 1985, Cincotta and Uresk 1986, Garret and Franklin 1988). Because of this small size, the resolution of the GIS map layers imposes an unrealistic decision-scale on the animals. The minimum pixel size had a resolution of one meter (park boundary and aerial photography); however, some databases had larger pixel sizes (5 to 30m). The cost surface is then based on a physical resolution determined by the databases and computational restrictions that do not match the dispersal capacity of a mid-sized rodent--prairie dogs cannot take 30 meter steps. Thus the ability to avoid barriers may be compromised with the model erring towards longer paths for linear features with high weights. By changing the weights of different classes of linear features (like roads). creating buffer zones around each linear feature (reducing edge effects), and mathematically enlarging features to match the pixel resolution, some of the path deviation is thought to be accounted for. However, some of the paths display jagged step-like movements, suggesting that some of the scale differences are still present.

Spatial Error--The spheroid and projection errors as they relate to precision are particularly important to mention for BADL. BADL straddles two UTM zones (13 and 14) and when dispersal paths are drawn across this junction (corrected for in GRASS GIS) any distortion due to the bending of the UTM zones at their edges is incorporated into the dispersal paths. Another source of spatial error involves the construction of the model itself through a series of vector and raster data transformations (Hunter and Goodchild 1995). The cost surface in GRASS is constructed as a raster surface, but it must be translated into a vector surface to calculate the dispersal path length. The model script does this in several steps: first the cost surface is corrected for map resolution, then 'drained' using a raster-based program to find the least-cost path. Once the path is constructed, it is translated into a vector format using a thinning algorthim that sees only 45 degree angles. The algorthim reduces curves to a series of small vectors, each at an angle of of 0, 45, 90, etc. Round features or long curves around features (like mesa tables and buttes) are more likely to have some degradation of area due to vector-averaging than long linear features (such as roads); however, there is no easy mechanism to quantify these differences.

Calculations of genetic distances--The Rogers' genetic distance used in this study has several underlying assumptions that can influence the metric relationship to physical distance (Jacquard 1970, Weir 1990, Weir and Cockerham 1990). Rogers' genetic distance is a metric measure and is often used in spatial studies as an analog for geographic distances (Stewart et al. 1992, White and Svenson 1992). Rogers' distance is more sensitive to disjunct or private alleles, while other measures of genetic distance use an overall allele frequency; therefore, short time frame studies are more appropriate with Rogers than those studies looking at speciation or phylogeny (Rogers 1991, Britten and Brussard 1991). Whether the twenty-one generational span used in this study is sufficient to overcome biases associated with private alleles or disjunct frequencies is not clear; however there is some evidence of private alleles in two of the populations used in this study (Bowser, unpublished data) which may contribute to the different groupings noted in the PCA analysis. Some statistics were tested using several other measures of genetic distance (Neis and the original Rogers); there were no significant differences in results produced by these statistics as compared to the Modified Rogers. Finally, if additional polymorphic alleles were discovered and added to the analysis, there may be some increased power in the genetic database and thus in the genetic distance grouping. A proposal has been submitted to repeat this study using micro-satellite DNA frequencies to calculate genetic distances for 1996 funding.

Discussion

The primary objective of this study was to demonstrate a new use for GIS technology by incorporating population genetics as a validation measure. Environmental features were shown to impact animal dispersal and thus create distances that were more similar to those predicted by the genetic measures. However, the overall model fit was poor suggesting that this exploration of GIS with genetics is still in its infancy and many bugs still need to be worked out. Dispersal paths, as a tool for understanding spatial relationships, may provide a key for understanding spatial relationships among populations on complex landscapes. Such models of animal movement and the impacts of such movement on long term population fitness are critical in the preservation of isolated species in preserves. As GIS develops into a management tool for the conservation of species, dispersal models can contribute to the realistic management of species--along with managing their genetic composition--on preserves into the next century.

This study used shell-scripts and UNIX to produce models; however, ideally these scripts should become tools portable to different ecosystems and species. As the park service is currently faced with several genetically isolated species, a model that looks directly at the impacts of different types of environmental features (including man-made features) on the genetics of a species can be immensely useful. I hope to expand this marriage of genetics and GIS to incorporate other databases, in new environments using different species--especially isolated or fragmented species. GIS, as an environmental modelling program, combined with genetics, as a measure of population dynamics, can be a powerful management tool and one with expanding application to preserves and parks.

Acknowledgements

I wish to thank my graduate advisors, Bette Loiselle and John Blake for their support and encouragement. The Technology Transfer Center of the National Biological Service provided assistance and equipment. Modelling advice and GIS expertise was generously provided by Ralph Root, Peter Strong, David Duran, and Susan Stitt of the NBS. Funding was provided by Dr. Bob Schiller of the National Park Service, Denver.

REFERENCES

Aquaah, G. 1990. Practical Protein Electrophoresis for Genetic Research. T. R. Dudley ed. Discorides Press, Portland Oregon.

Aspinall, R. and N. Veitch 1993. Habitat mapping from satellite imagery and wildlife survey data using a bayesian modeling procedure in a GIS. Photogrammetric Engineering and Remote Sensing 59(4): 537-543.

Awaida and J. Westervelt 1993 r.cost Manual page. GRASS 4.1 User ReferenceManual 1993 USACerl publication

Chesser, R. K. 1983. Genetic variability within and among populations of the black-tailed prairie dog. Evolution 37(2): 320-331.

Cincotta, R. P. and D. W. Uresk. 1986. Ecology of The Black-tailed Prairie Dog in The Badlands National Park. National Forest Service Research Experimental Station report. Rapid City, South Dakota.

Daley. J. G. 1992. Population reductions and genetic variability in black-tailed prairie dogs. Journal of Wildife Management 56(2): 212-220.

Dunning, J. B., B. J. Danielson and H. R. Pulliman 1992. Ecological processes that affect populations in complex landscapes. Oikos 65: 169-175.

Goodchild. M. 1993 Introduction in Environmental Modelling with GIS. Oxford University Press London

Green, K., D. Kempka, and L. Lackey 1994. Using remote sensing to detect and monitor land-cover and land-use change. Photogrammetric Engineering and Remote Sensing 60(3): 331-337.

Hartl, D .L. and A. G. Clark 1989. Principals of Population Genetics. Sinauer Associates Sunderlands, MA.

Hillis, D. M. and C. Mortiz 1990. Molecular Systematics. Sinauer Associates, MA.

Hunter, G. J. and M. F. Goodchild 1995. Dealing with error in spatial databases: a simple case study. Photogrammetric Engineering and Remote Sensing 61(5): 529-537

Jacquard, A. 1970. The Genetic Structure of Populations Springer-Verlag New York

Johnson, L. B. 1990. Analysis of spatial and temporal phenomena using geographical information systems. Landscape Ecology 4(1): 31-43.

Johnston, C. A. and R. J. Naiman 1990. The use of a geographic information system to analyze long-term landscape alteration by beaver. Landscape Ecology 4: 5-19.

Kareiva, P. 1990. Population dynamics in spatially complex environments: theory and data. Philosophical Transactions of the Royal Society, London B. 330 Pages: 175-190.

Keller, J. K. 1990. Using aerial photography to model species-habitat relationships: the importance of habitat size and shape. pp 34-46. in Mitchell, R. S., C. J. Sheviak, and D. J. Leopold eds. Ecosystem Mangement: Rare species and significant habitats. vol bull 471. New York, New York State Museum.

Kienast, F. 1993. Analysis of historic landscape patterns with Geographic Information Systems (GIS)--a methodological outline. Landscape Ecology 8(2):103-118.

Kolasa, J. and C. D. Rollo 1991 Introduction: the heterogeneity of heterogeneity: a glossary. in Ecological Heterogeneity J. Kolasa and S.T.A. Pickett eds. Springer-Verlag New York.

Kozakiewicz. 1993. Habitat isolation and ecological barriersthe effect on small mammal populations and communities. Acta Theriologia 38: 130.

Lande, R. 1988. Genetics and demography in biological conservation. Science 241: 1455-1460.

Leberg, P. L. 1992. Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. Evolution 46: 477-494.

Mitchell-Olds, T. 1992. Does environmental variation maintain genetic variation? A question of scale. TREE 7: 397-398.

Opdam, P. 1991. Metapopulation theory and habitat fragmentation: a review of holartic breeding bird studies. Landscape Ecology 5: 93-106.

Pasteur, N., G. Pasteur, F. Bonhome, J. Catalan, and J. Britton-Davidson 1988. Practical Isozyme Genetics. Ellis Horwood Limited, Chicester, West Sussex, England.

Richardson, B. J., P. R. Baverstock, and M. Adams 1986 Allozyme Electrophoresis: a handbook for animal systematics and population studies. Academic Press, London.

Rogers, J. S. 1991. A comparision of the suitability of the Rogers, Modified Rogers, Manhatten, Cavalli-Sforza and Edwards distances for inferring phylogenetic trees from allele frequencies. Systemmatic Zoology 40: 63-73

Schenbeck, G. L. and R. J. Myhre 1986. Aerial photography for assessment of black-tailed prairie dog management on the Buffalo Gap National Grassland, South Dakota. USDA-FS report 86-7.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T.C. Edwards jr. J. Ulliman, and R.G. Wright. 1993. GAP analysis: a geographic approach to protection of biological diversity. Wildlife Monographs 123.

Slatkin, M. 1981. Estimating levels of gene flow in natural populations. Genetics 99: 323-335

------------ 1985 Rare alleles as indicators of gene flow. Evolution 39:53-65.

------------ 1987 Gene flow and the geographic structure of natural populations: Science 236: 787-792.

----------- 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47: 264-279

Stewart, D. T. , A. J. Baker and S. P. Hindocha. 1993. Genetic differentiation and population structure in *Sorex haydeni* and *S. cinereus*. Journal of Mammalogy 74(1): 21-32.

Swofford, D. L. and R. B. Selander 1981. BIOSYS-1. A FORTRAN program for the comprehensive analysis of electrophorectic data in population genetics. Journal of Heredity 72:281-283

Turner, M. G. and R. H. Gardner. 1991. Quantitative Methods in Landscape Ecology. SpringerVerlag, New York.

Waggoner, G. 1989. Analysis of alternative road alignments using GRASS 3.0. in Proceedings of the 1988 Geographical Resource Analysis Support System (GRASS) User Group Meeting. USACERL Technical Manuscript N-89/18 September 1989.

Wayne, R. K., S. B. George, D. Gilbert, P. W. Collins, S. D. Kovach, D. Girman, N. Lehman. 1991. A morphological and genetic study of the island fox *Urocyon littoralis*. Evolution 45:1849-1868.

Weir 1990. Genetic Data Analysis. Sinaur Associates, Sunderland MA.

Weir, B. S. and C. C. Cockerham 1984. Estimating F-statistics for the analysis of population structure. Evolution 46: 608-615.

White, M. N. and G. E. Svenson. 1992. Spatial-genetic structure in the eastern chipmunk Tamias striatus. Journal of Mammalogy 73:619-624.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395-420.

Author Information

Gillian Bowser, Natural Resource Specialist, Rocky Mountain System Support Office--National Park Service, 12795 W. Alameda Parkway, Denver, Colorado 80225.

phone: 303-969-2865

fax: 303-969-2644

email: gillian@rmro.nps.gov

url: http://www.nps.gov/wildlife/nat.resources/gillian/

FRED BOGS1, JAMES NEWELL1, AND JEFFREY FITZGERALD2

1 Institute of Applied Sciences, University of North Texas, Denton, Texas 76203 USA
2 Center for Spatial Analysis and Mapping, University of North Texas, Denton, Texas 76203 USA

MODELING SPATIAL AFFECTS OF LANDSCAPE PATTERN ON THE SPREAD OF AIRBORNE FUNGAL DISEASE IN SIMULATED AGRICULTURAL LANDSCAPES

Abstract The progression of an airborne fungal disease within an agricultural landscape is an important ecological process that is influenced by landscape pattern. Plant epidemiology research has led to the creation of several regression-based probabilistic models of disease-focus expansion. These models have no explicit spatial component and do not account for heterogeneity at the site of investigation. To better understand the relationship between the pattern of an agricultural landscape and the spread of airborne fungal disease, a raster-based GIS simulation model was developed and implemented in Arc/Info(R) GRID. The model treats infection dynamics as a diffusion process and vegetation as spore-filtering media. The GIS-based model allowed for investigation of the affect of variations in vegetation patterns on disease spread. Using the GIS-based model, a two-crop experiment was conducted. Disease spread was simulated for one-hectare fields with varying arrangements of wheat and maize or fallow. Landscape metrics sensitive to boundary shapes and fragmentation were calculated for each crop raster. Increasing connectivity of wheat resulted in increased infection in the maize simulations. Introducing narrow fallow patches to homogeneous wheat fields increased spread.

INTRODUCTION

The patterns of landscape development in time and space result from complex interactions of physical, biological, and social forces (Risser 1984). The alteration of natural land cover patterns results in a mixture of natural and human-managed patches. The size-shape relationships of the altered land cover can influence a number of important ecological phenomena (Beasley 1981, Burgess 1981), including migration, speciation, and the spread of disturbance (Turner 1987). Gradients of disturbance frequency and severity are often controlled by physical or vegetational features.

Goals and Objectives

The goal of this research has been to explore relationships between landscape pattern and the spread of disease by modeling the spatial and temporal development of an airborne fungal disease in varying patterns of susceptible and non-susceptible plant communities. Specifically, our objective was the formulation and GIS implementation of a spatial allocation model for airborne crop disease. Simulations were performed to study the response of fungal crop disease to small-area landscape pattern. Statistical analyses of this relationship were performed. An additional objective was to assess the usefulness and applicability of an off-line response surface design which would allow parameter searches and optimization techniques to be applied to simulation results. Response surface analysis is proposed as a method to minimize the cost of finding target landscape configurations that will minimize transmission of infection.

Rationale

Heterogeneity of the landscape may enhance or retard the spread of disturbance (Turner 1987). The incidence of diseases is related to the spatial arrangement of crops and cultivars (Perrin 1980). According to Alexander (1989), spatial heterogeneity, as it applies to phytopathology, is the variation in plant density, plant genetic composition, and plant location with respect to the physical environment. A plant pathogen may be readily transported over long distances and placed in an uninvaded area, but unless it can establish and maintain itself there, its geographical range will not be increased (Roberts 1975). Murdoch (1975) postulated that it may be possible to design agroecosystems that minimize pest problems and reduce the need for active control measures. Landscape connectivity may be important for the persistence of organisms (Turner 1989, Wegner 1979). Therefore, by fragmenting or structuring the landscape to minimize connectivity, the spread of disturbance may be minimized. Alexander (1989) added that "by mimicking aspects of these natural populations and communities, it was postulated that disease control could be achieved in more natural, long-lasting ways." It is axiomatic that plant pathogens cannot establish themselves in new areas unless susceptible plants occur there. Barriers in the form of fallow or non-susceptible species may therefore be important in retarding the spread of an airborne disease. The dispersal of pests may be impeded where host and non-host grow together (Perrin 1980).

Understanding the role of landscape pattern in the spread of disturbance is fundamental to the development of functional landscape management scenarios. GIS-based modeling incorporates the spatial component of disease spread, allowing spatial scales and patterns to be easily altered, and allowing flexibility in modeling different cover types, disturbances, spatial patterns, and weather conditions.

Background

Phytopathological modeling predicts infection probability or severity within a stand of vegetation given that the severity of a source is known at a point in time. Severity is measured as lesion density or spore coverage per unit area. For fungal rusts and blights, probability of infection and infection severity depend on spore dispersal mechanisms. Methods for modeling the spread of fungal plant disease approach the problem from different perspectives: (1) a mechanistic approach, (2) gradient modeling, (3) stochastic simulation, and (4) spatio-temporal autoregression modeling.

Aylor (1986) proposed a mechanistic model for regional spread of uredeosproes over distances on the order of 1000 km. Probability of infection at a site is based on a rigorous accounting of spores from production through deposition onto the crop. This is a compartmental model that provides a useful framework for understanding the local processes of infection spread (e.g., spore release and transport by wind). Some processes, such as spore loss to UV-radiation, that are significant at regional scales do not influence local dispersion. A disadvantage of the mechanistic approach for simulation modeling is the expense of obtaining precise spore counts at extreme concentrations. Furthermore, a high degree of computational precision is required to implement a complex mechanistic model which incorporates meteorological factors.

Gradient models such as the Gregory model (Gregory 1968, Mundt 1989) use an epidemiological gradient to express the relationship of infection severity to distance. The relationship is usually linearized by applying a logarithmic transformation to the measurement of distance from an infection source and a gompit, probit, or corrected logarithmic transformation to the severity metric. The power of the gradient breaks down with increasing complexity of terrain topology and interacting infection sources. Furthermore, predicting severity a set distance from a source requires intensive spore inventory, and extrapolation from off-center points is not robust (Headrick 1988).

Minogue (1989) advises stochastic simulation of the spread of infection based on sampling from a realistic distribution of dispersal distances. This is a spore-conservative approach. Distributions that result in a wavelike diffusion pattern are to be avoided. Infinite-mean distributions such as the Pareto distribution tend to produce new foci and result in more realistic patterns of dispersal than do finite-mean distributions such as the exponential distribution. Statistical spore-distance distributions resemble the shapes of gradient curves. In fact, the Pareto distribution and the Gregory model have the same form, $y = ax^b$. In order to be practical for simulation, the spore map should be aggregated into raster cells. Otherwise, tables of disease lesions necessary for computing release at time t+1 become enormous.

The most easily calibrated models are the spatio-temporal autoregression models (Reynolds 1988). A linear combination of distance- and time-lag terms is used to compute the severity or infection probability at a point within a stand of vegetation. Different linear time-series functions can be fit to along-row and across-row spread. These methods have the advantage of applying statistically legitimate techniques to observations with time-autocorrelated error terms. The rigid treatment of spatial configuration causes these models to be less suitable than other techniques for theoretical simulations of disease spread in heterogeneous landscapes.

New lesions formed on a single plant, on adjacent plants, and on plants some distance away are the result of common dispersion processes. The simulation model described herein partitions the increase of disease severity into two processes: severity increase within a cell and the spread to other cells. The logistic growth equation was used to model the increase in disease severity, while a stochastic model was used to model the transmission of infection between cells. An all-or-none distinction is made between infected and uninfected raster cells. A cell-based model is used to reduce computational requirements.

METHODS

Model

Crop disease was chosen as an ecological disturbance whose spatio-temporal progress could be modeled in a straightforward manner. The model was chosen and formulated so that the effect of patches of non-susceptible crops and buffers of unplanted space on disease spread could be investigated. Landscape topology had to be communicated in a modular manner so that patterns could be varied without changing the simulation engine.

A raster-based data structure for stochastic simulation of local spore dispersal mechanisms was devised. The rationale for this approach were that: (1) the analytical models describing the advance of an infection frontier using regressions between logarithmic transformations of distance and disease severity (Headrick 1988, Mundt 1989) are spatially too general for exploration of various landscape patterns; and (2) at the landscape scale, ideal mixture models are functionally homogeneous (Bosch 1990). The equations used to predict rates of focus expansion in homogeneous landscapes are not applicable to heterogeneous sites.

Disease propagation was partitioned into three sources, exposure, infection, and severity. Exposure is the event that the number of spores accumulating in a cell exceeds an arbitrary threshold. Infection occurs when a cell becomes a host to the disease and is able to expose other cells. Severity is a measure of the amount of infection within a cell.

Our model is a simple stochastic simulation model based on the deterministic probabilities of airborne spore dispersal. It partitions disease severity into two proscesses: (1) severity increase within a cell (modeled by a logistic growth equation), and (2) the spread to other cells (modeled stochastically).

The simulation model must generate probabilities of *exposure* and *infection* at each timestep over the entire crop raster. The spatial model was adapted from Minogoue's (1989) model of disease spread as a diffusion of spores through a filter of vegetation. The probability of exposure drops off exponentially with distance. The gradient of disease spread depends on the media (i.e., vegetation height and density) through which it is filtered. The input crop rasters were coded so that each cover type had associated with it a "distance to 50% probability of infection." This spore-filtering capacity is analogous to the directional gradient of the Gregory (1968) model, but applies to both susceptible and non-susceptible vegetation.

Exposure A raster cell becomes infected during timestep *t* if both exposure and infection occur within the cell. The event that both exposure and infection occur is affected by the susceptibility of the crop within a cell, distance to neighboring infected cells, and the severity of infection in neighboring cells. Susceptibility is defined as the probability of infection given exposure. If we let *(i,j)* be the coordinates of an uninfected susceptible raster cell and *(m,n)* be the coordinates of the infected cell nearest *(i,j)*, the probability that cell *(i,j)* becomes infected is then given by

$$P_{i,j}[\mathit{infection} \cap \mathit{exposure}] = P_{i,j}[\mathit{infection} \mid \mathit{exposure}] \times P_{i,j}[\mathit{exposure}] \quad (1)$$

or

$$I_{i,j} = S_{i,j} \times E_{i,j} \quad (2)$$

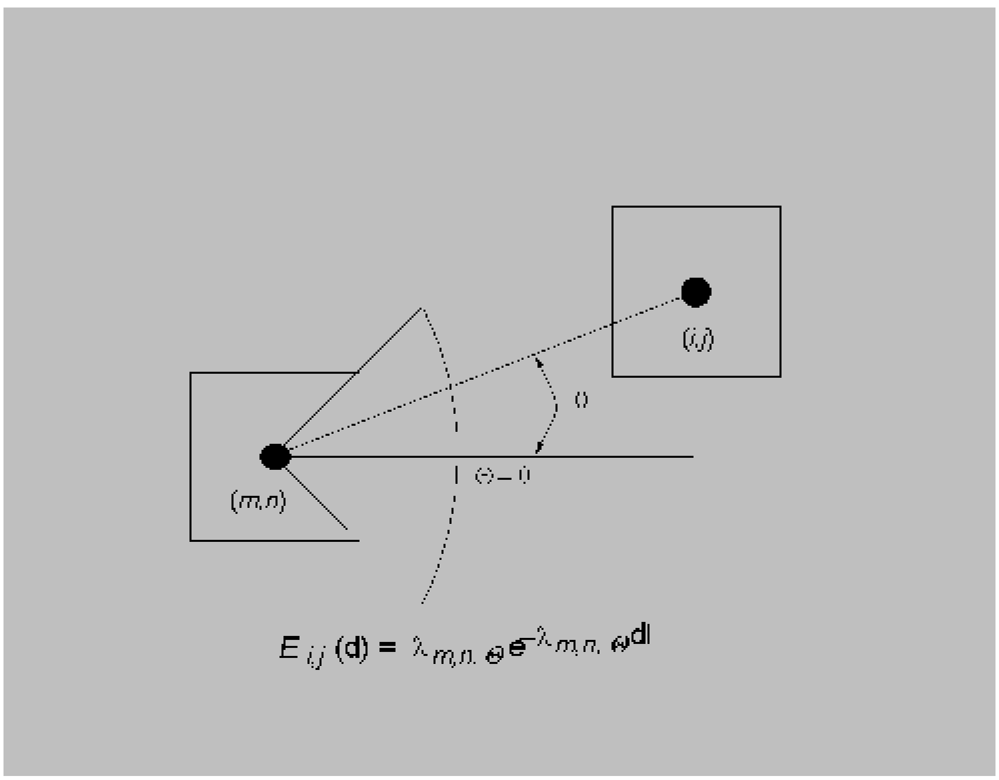
where × denotes scalar multiplication. Susceptibility is a random variable. It has been considered a constant parameter for a given crop.

The simulation model must generate probabilities for exposure and infection events at each timestep over the entire crop raster. The spatial model was adapted from a treatment of disease spread modeled as diffusion of spores through a filter of vegetation (Minogue 1989). The probability E of sufficient exposure at *(i,j)* by *(m,n)* is given by

$$E_{i,j}(d) = \lambda_{m,n,\Theta} e^{-\lambda_{m,n,\Theta} d} \quad (3)$$

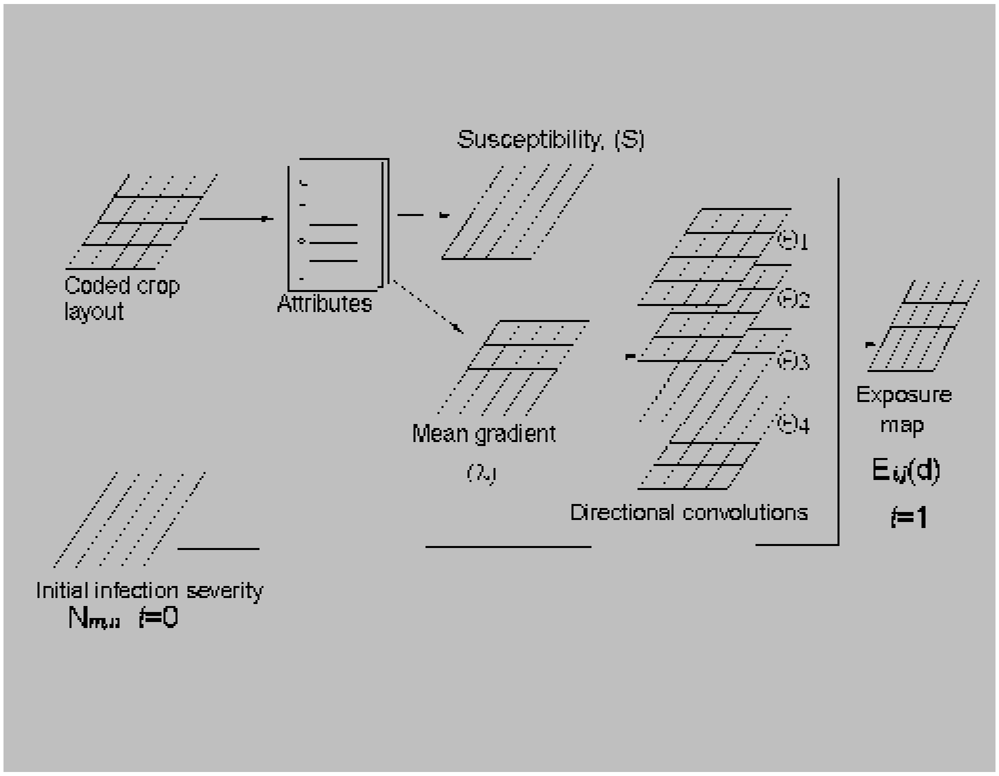
where d is the Euclidean distance between *(i,j)* and *(m,j)* and $\lambda_{m,n,\Theta}$ is the decay constant for direction class Θ between *(i,j)* and *(m,n)* (Figure 1). Angles θ were classified into wedges of $\frac{\pi}{2}$ centered on $\Theta \times \frac{\pi}{2}$ $(\Theta \text{ in } 1,2,3,4)$ (Figure 2). Parameter λ is a crop-cover-specific parameter related to the spore-filtering capacity of the intervening vegetation between a host and a potential host, which will vary with plant height and foliage density. It is analogous to the directional gradient of the Gregory model (Gregory 1968), but applies to susceptible and nonsusceptible vegetation. No differentiation was made in broadcast capabilities for different plants resulting from different heights and foliage densities.

Figure 1. Probability of exposure among cells.



click on image for higher resolution

Figure 2. Input preprocessing: steps necessary for calculation of exposure.



click on image for higher resolution

Severity Based on the assumption that a host's propensity to expose nearby potential hosts is related to its severity, the intercept of equation (3) is adjusted to include a response to severity:

$$E_{i,j}(d) = I_{m,n} \times \lambda_{m,n} \times e^{-\lambda_{m,n,\Theta} d} \qquad (4)$$

where $I_{m,n}$ is defined on [0,1) and calculated as

$$I_{m,n} = G + N_{m,n,t}(1 - G) \qquad (5)$$

where $N_{m,n,t}$ gives the severity at (m,n) and ranges over [0,k), where *k* gives the maximum severity a susceptible cell can sustain. G is the minimum adjustment, a weak global parameter. A smooth, monotonically increasing function converging to 1 could have been used but was avoided to save computation time at this step.

The progress of disease severity through time at a fixed location was modeled by the discrete-time logistic growth equation

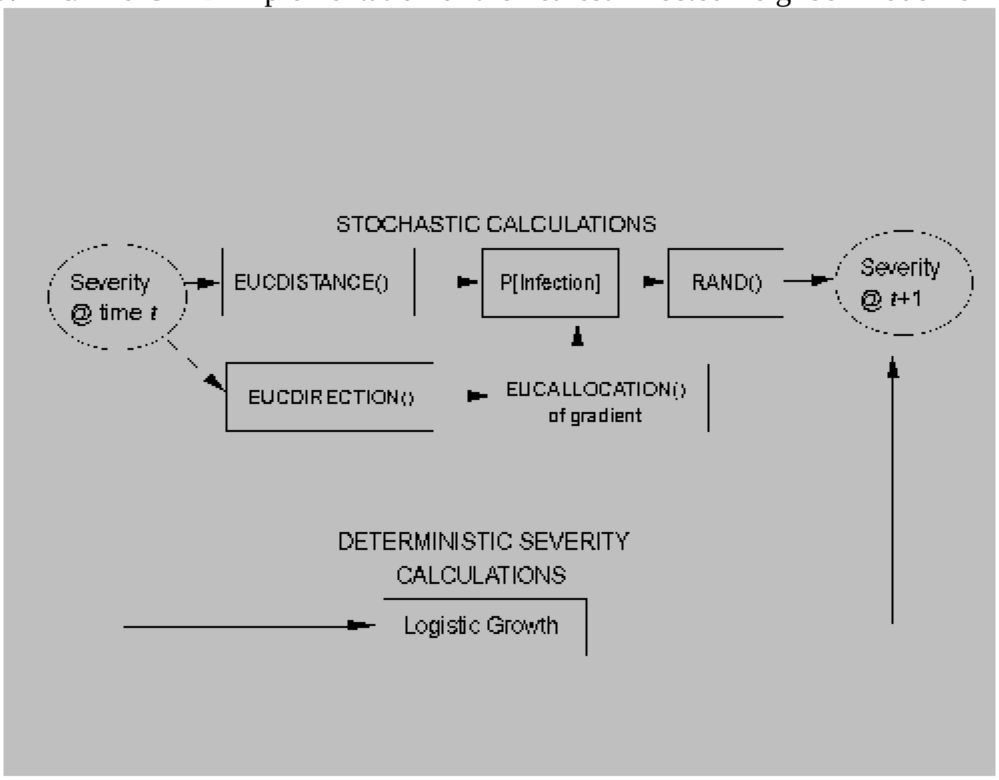
$$N_{m,n,t} = N_{m,n,t-1} + r(1 - N_{m,n,t-1}/k),\ t \text{ in } [1, 2, 3, \ldots] \qquad (6)$$

where *r* represents an intrinsic growth rate. The parameter *r* was defined as a constant in this simulation but may depend on environmental effects such as plant health, temperature, and moisture which vary through time.

Model Implementation

The GRID module of Arc/Info(R) allows for the creation of input crop-cover maps and formula-based implementation of the model. Spatial interactions within the model were accomplished using GRID's cost-allocation functions. It was required that the distance and angle between *(i,j)* and *(m,n)* be calculated and that the parameter $\lambda_{m,n}$ be allocated to *(i,j)* at runtime. A diagram of the processing structure of the simulation model is given in Figure 3.

Figure 3. Arc/Info GRID implementation of the nearest-infected-neighbor model for local spread.



click on image for higher resolution

Let *C* be a matrix representing the arrangement of crops. Let *c* represent a particular crop type. Let μ_c be the mean of the exponential distribution in equation (3) which characterizes the spore-filtering capability of crop *c*. Let *M* be a matrix of parameters μ_c corresponding to *C*.

In order to preprocess the input raster *C*, raster *M* must be generated by merging raster *C* with a database of parameters μ_c. From *M*, four rasters, $\bar{M}_\Theta$ $(\Theta \text{ in } \{0, 1, 2, 3\})$, must be generated. These $\bar{M}_\Theta$ are computed by smoothing *M* with wedge-shaped convolution window W_Θ. W_Θ has central angle $\Theta \times \frac{\pi}{2}$ and is populated by

$$W_{i,j} = \frac{e^{-\kappa \delta_{p,q}}}{\sum e^{-\kappa \delta_{i,j}}}, \text{ where} \tag{7}$$

$$\delta \| \vec{p,q} \| \leq \rho$$

$$\Theta - \frac{\pi}{4} \leq atan(\frac{p}{q}) < \Theta + \frac{\pi}{4}$$

p, q in $[1, 2, 3, ...]$

Parameter ρ is the radius of the wedge (5m used) and κ determines the distance decay of interactivity. The parameter $\lambda_{i,j,\Theta}$ is derived from $\overline{M}_{i,j,\Theta}$ at runtime from the expression

$$\lambda_{i,j,\Theta} = \frac{1}{\overline{M}_{i,j,\Theta}} \quad (8)$$

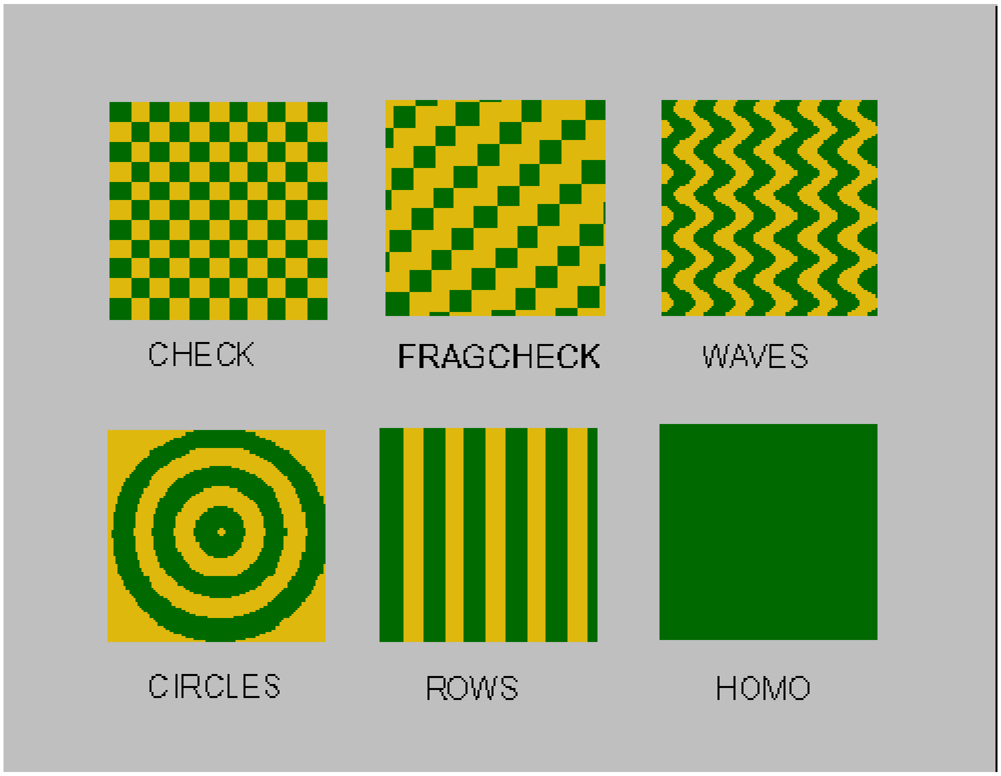
which holds since μ_r is the mean of an exponential distribution.

Since GRID's spatial interaction functions EUCALLOCATION, EUCDISTANCE, and EUCDIRECTION take integer rasters as arguments, an internal magnification of was used. To reduce calculations at a timestep, a cutoff value of 10m was applied to the spatial interaction modeling as suggested by Minogue (1989). Given , the 10m cutoff value results in discarding only of the probability.

Simulations

Three replicates of the homogeneous wheat simulation and five simulations each of the input patterns, varying barrier width and type, were conducted. The homogeneous simulations, HOMO, were used as zero barrier width patterns. Fallow patches model the situation where there is little or no obstruction to the spore cloud. Maize patches represent a barrier that is taller and more dense than wheat. In order to determine barrier width, a test strip of alternating bands of susceptible crop and increasing barrier widths was designed. The model was run until no further spread was observed. Spread ceased at barrier width of 10m in fallow and 7m in maize. Therefore, barrier widths of 3m, 6m, and 9m were chosen for the simulations. Rasters containing five patterns (ROWS, WAVES, CIRCLES, CHECK, and FRAGCHECK) were input to the model with barrier widths of 3, 6, and 9 meters (Figure 4). Barrier widths were the length in the x-direction of the barrier patches. FRAGCHECK has less diagonal connectivity than CHECK. ROWS are linear swaths of alternating susceptible and nonsusceptible cover. WAVES is a sinusoidal variation of ROWS. CIRCLES consists of closed annuli arranged concentrically about the center of the crop raster. Simulation rasters were 100 cells on a side, or one hectare. Generation of the patterns was facilitated by a graphical user interface (GUI) to algorithms developed in the Arc/Info AML script language. The homogeneous simulations were considered the zero barrier width point for each pattern type.

Figure 4. Crop patterns input to the simulation model.



Simulations were conducted for 10 timesteps of equal length. Infection foci were a single cell near the center of each raster. For the CIRCLES pattern, infections were started in outer annuli because it was undesirable to confine infection to the center circle, a patch atypical of the overall pattern.

Pattern Quantification

Landscape pattern was quantified so that is influence on local airborne crop disease mechanisms could be evaluated. Input patterns were analyzed using Landstat descriptive landscape statistics software (Riitters 1994). Perimeter/area ratio was used as a simple measure of shape complexity. As the complexity of the landscape increases, the ratio also increases.

Response Surface Design

Optimal response surface designs for resolving crossed effects in mixtures have design points at vertices of a regular k-dimensional simplex, a geometric figure with facets created by intersecting hyperplanes. k represents the dimension of the model or number of independent variables. Mixture designs, having all positive component concentrations bounded by a ceiling of 100% or less are confined to the surface of a figure no larger than a k-simplex. Block designs can be combined with mixture designs to allow process variables into the model.

A response surface was fit to the data to analyze the dependency of areal spread after 10 timesteps on the landscape pattern metrics. The model form is derived by Khuri and Cornell (1987), and combines process variables and mixture component variables. The model fit to the data was composed of linear and crossproduct terms only, with no intercept term. To reduce the number of terms in the model, a stepwise algorithm was applied. Standardized parameter coefficients were computed by dividing the estimate for each parameter by the ratio of the independent parameter standard error to the standard deviation of the dependent variable. Standardized parameter coefficients were used as unitless estimators of the effect of the landscape metrics on the response variable. Lack-of-fit tests were applied to response surfaces in an iterative manner, with order increasing until the hypothesis of zero lack-of-fit could not be rejected at 95% significance (following Khuri and Cornell 1987).

RESULTS

One simulation run for each variation (i.e., changing barrier width and type) of the CHECK pattern is illustrated in figure 5. The simulation shown in figure 5 was repeated for each pattern. To compare patterns, a normalized infected area (NIA) was determined for each pattern. NIA was calculated as total infected area divided by total susceptible area. The NIA was calculated for each pattern-barrier-width combination. Because the simulations were stochastic mean NIA was determined for each pattern. The HOMO and CHECK matrices are used to demonstrate the range of variance associated with the stochastic simulations (Figure 6). Bartlett's test applied to infected area of the landscape patterns suggested that variances were not equal among the maize and fallow observations (P= 0.95). Mean NIA, for both fallow and maize barriers, and P/A for the five landscape patterns are given in figures 7 and 8.

Figure 5. Infection severity after 10 timesteps.

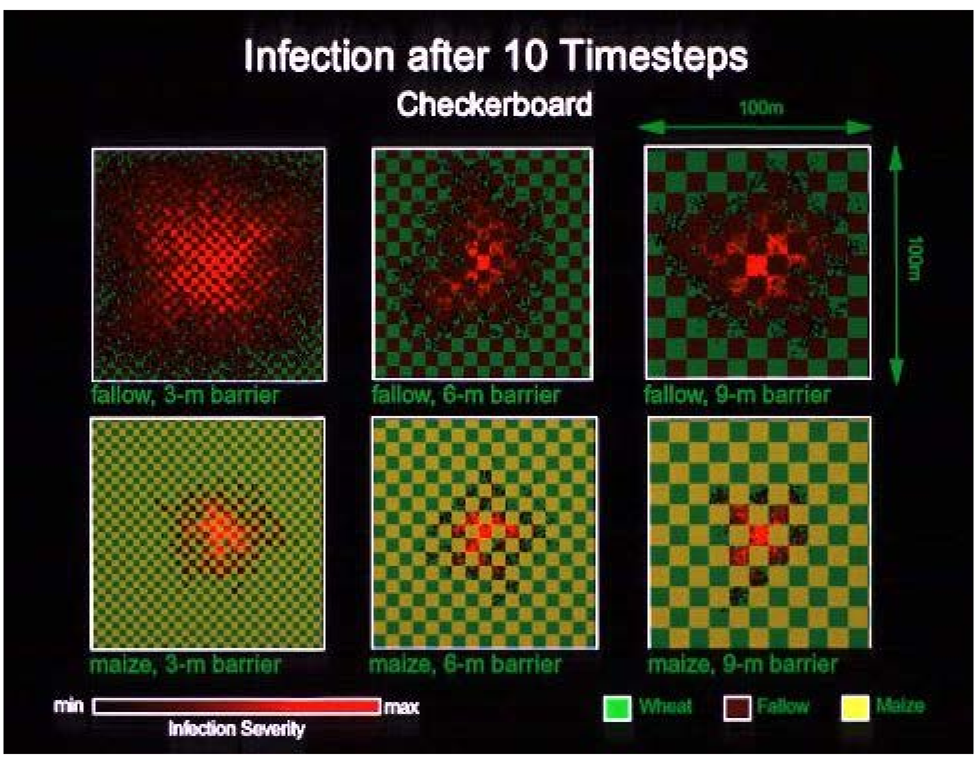
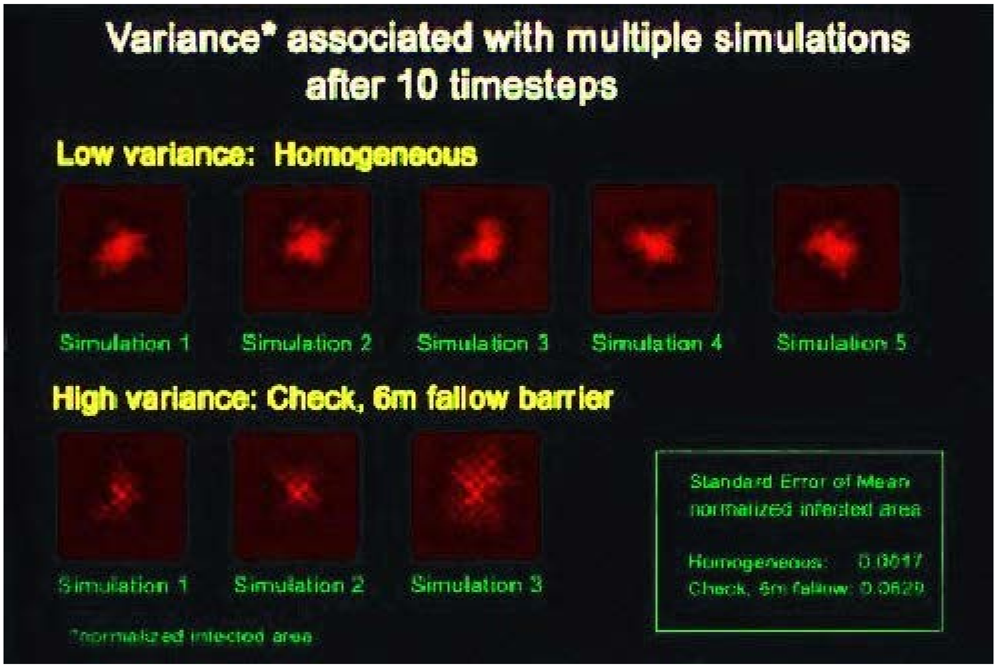


Figure 6. Variance associated with multiple simulations after 10 timesteps.



Fallow Barrier

The mean NIA after 10 time steps for the fallow simulations was 30.1% of total susceptible area, with a standard deviation of 16.3%. The fallow infection data were normally distributed (Shapiro-Wilk W= 0.8, $p > 0.05$). NIA for single fallow runs ranged from 4.5% to 80% of susceptible cover. The maximum mean of three replicates for this measurement was observed in the CHECK pattern with 3m-fallow barrier; the minimum mean infected area occurred in the ROW-pattern with 9m barrier. For CHECK, a trifold increase in NIA over homogeneous is seen in the 3m-fallow barrier runs but, NIA the 6m and 9m fallow barriers was less than for HOMO. For all patterns except ROWS, the NIA decreased with increased barrier width. The 3m-fallow barrier for the ROWS pattern had lower NIA than did the 6m-fallow barrier, with NIA for the 9m-fallow barrier less than the 3m-fallow barrier (Figure 8).

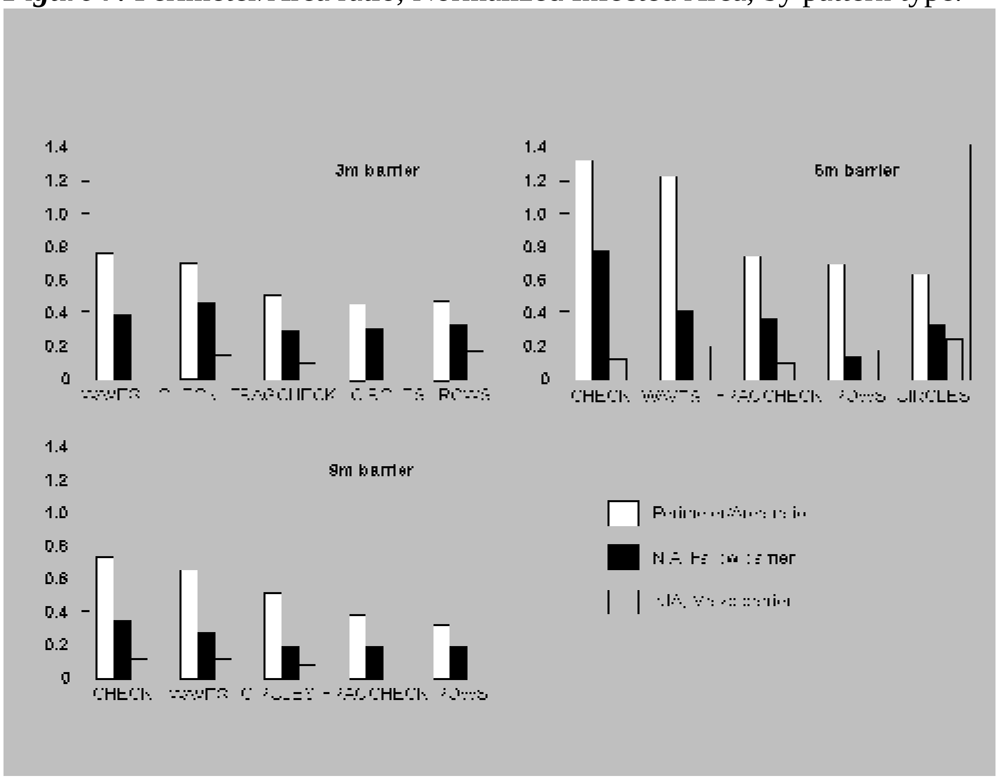
Maize Barrier

The mean NIA was lower for the maize barriers than for the fallow barriers, with mean and standard deviation of 12.5% and 5.3%, respectively. The maize data were normally distributed (Shapiro-Wilk W= 0.937, $p < 0.05$). Single observations of NIA for the maize barrier simulations ranged from 4.5% to 80.5%. Maximum and minimum means were observed in the homogeneous simulations and the 3m-maize CHECK patterns, respectively. The maize CIRCLES, FRAGCHECK, ROWS, and WAVES patterns had a decrease in areal spread with increasing barrier width. The exception was CHECK, which had an increase in NIA with barrier width. The range of NIA for CHECK was small and may represent no change with barrier width (Figure 8).

Pattern Quantification

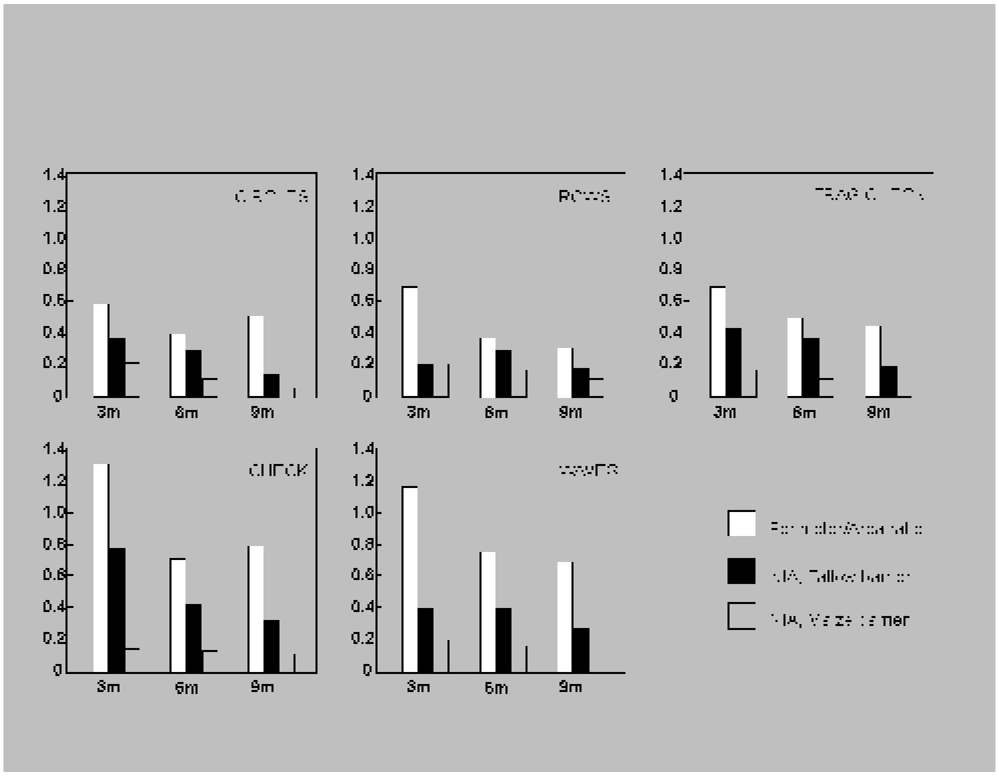
Pattern metrics (perimeter to area, P/A, ratios) for resource patches were used to quantify fragmentation and to compare simulation results (Figures 7 and 8). Pattern complexity (i.e., resource patch fragmentation) increases as NIA decreases. For all three barrier widths, CHECK and WAVE had the highest P/A ratios. ROWS and CIRCLE had the lowest P/A ratios, with FRAGCHECK having intermediate values. Comparison of absolute P/A values is complicated by the influence of raster boundaries on the area of resource patches. For example, P/A for CIRCLE grid with 9m barrier is greater than P/A in the 6m barrier. This results from the number of resource patches at the edge.

Figure 7. Perimeter/Area ratio, Normalized Infected Area, by pattern type.



click on image for higher resolution

Figure 8. Perimeter/Area ratio, Normalized Infected Area, by barrier width.



click on image for higher resolution

DISCUSSION

Pattern-Disease Spread Relationships

As shown by the plots of normalized infection spread (Figures 7 and 8), the fallow barrier seems to exhibit a dual function. Greater areal spread was observed in all three simulations of the 3m fallow CHECK pattern than in any of the homogeneous simulations (Figure 5). This may be attributed to the reduced filtering capacity of fallow as opposed to wheat. A spore is transmitted more readily through the fallow than wheat barrier. At the 3m barrier width, there is greater potential for dramatic spread due to the small patches of open space.

Contiguous susceptible crop favors the spread of infection until local saturation of the vegetation with disease spores occurs. At saturation, spores released within the patch are largely "wasted" by landing on infected vegetation. When barriers become wide enough, however, infection becomes lower than in homogeneous simulations. The low mean areal spread observed for the 3m ROWS pattern is difficult to explain. The maximum spread observed for WAVES at 6m results from the normalization of susceptible area. A monotonically decreasing trend is seen in the raw WAVES infection data. For the maize WAVES runs, maximum spread also occurs for 6m barriers in the normalized, but not in the raw data. The decrease in normalized areal spread for the other patterns at 6m and 9m barrier widths suggests that interruption of the susceptible crop with large areas of bare patches or swaths of nonsusceptible crops reduces spread by local mechanisms.

Barrier width and amount of susceptible crop are positively correlated. The CHECK pattern is exceptional in that barrier cells always occupy half the raster, with the other half allocated to susceptible crop. For the fallow scenario, a trifold increase in normalized spread is seen in the 3m run relative to the homogeneous run. Adding connected patches of a barrier, such as maize, lessens the areal spread of infection. An increase in normalized spread with barrier width is evident in the 3m, 6m, and 9m runs of CHECK. Only for the CHECK pattern is there a positive relationship between barrier width and connectivity among susceptible patches. For the other patterns, increasing the barrier size increases the fragmentation of the susceptible areas.

Response Surface

Against Khuri and Cornell's (1987) criteria, our simulation design was suboptimal for a response surface analysis. Two caveats apply when the response model is suspected of being unstable due to imperfect design: (1) response estimates are not robust outside the boundaries of the design points and (2) the parameter estimates may not resolve the effects of interaction among terms. The model may fit the design points well, but the relationships indicated by the parameter estimates may not apply in general. Searches for optimal parameter combinations were forgone in this phase of the study because of the unbalanced design.

The work presented in this study could be used to design a more robust response surface analysis. A variable transformation which equalizes variance in the areal infection parameter would increase the power of response model significance testing. It is important to take advantage of what is known about the behavior of the variance when formulating an experimental design so that the power of statistical tests for model significance can be optimized.

Ongoing and Future Research

A simple disease dispersion model was implemented on a GIS platform. Rigorous validation of this model was not conducted based on the assumption that broad generalizations could be made using a model which captured the essential nature of biological processes. Calibration is necessary in order to prescribe landscape configurations that contain or impede the spread of airborne rust.

Further study is required to refine and execute the research designs described herein. The model was designed to be flexible enough that important environmental factors (e.g., wind, moisture, topography) could be incorporated. A Pareto distribution to model the distance-probabilities of exposure may help reduce any wave-like properties of simulated disease spread. The model should be refined and calibrated with measured field data, including integrating more realistic landscape patterns and environmental factors. We have shown that GIS-based simulation models can be used to explore the relationship between landscape pattern and ecological disturbance.

This paper and research is dedicated to the memory of Fred Bogs. He was a bright and talented young student who is deeply missed by all who knew and worked with him.

LITERATURE CITED

Alexander, H.M. (1989) Spatial heterogeneity and disease in natural populations. *in*, *Spatial Components of Plant Disease Epidemics*, ed. Michael Jeger, New Jersey: Prentice-Hall, Inc. pp. 144-164.

Aylor, D.E. (1986) A framework for examining inter-regional aerial transport of fungal spores. *Agricultural and Forest Meteorology* 38:263-288.

Beasely, D.B. and L.F. Huggins. (1981) *Answers Users Manual*. EPA-905/9-82-001. U.S. Environmental Protection Agency, Region V, Chicago, Ill.

Bosch, F. van den, M.A. Verhaar, A.A. Buiel, W. Hoogkamer, and J.C. Zadoks. (1990) Focus expansion in plant disease IV: Expansion Rates in mixtures of resistant and susceptible hosts. *Phytopathology* 80:598-602.

Box, G.E.P., and N.R. Draper. (1987) *Empirical model-building and response surfaces*. New York: John Wiley and Sons.

Burgess, R.L., and D.M. Sharpe, eds. (1981) *Forest island systems in man-dominated landscapes*. New York: Springer-Verlag.

Gregory, P.H. (1968) Interpreting plant disease dispersal gradients. *Annual Review of Phytopathology* 6:189-212.

Headrick, J.M., and J.K. Pataky, (1988) Spatial and temporal development of common rust in susceptible and partially resistant sweet corn hybrids. *Phytopathology* 78(2):227-233.

Khuri, A.I., and J.A. Cornell. (1987) Response surfaces: design and analyses. New York: Marcel Dekker.

Minogue, K.P. (1989) Diffusion and Spatial Probability Models for Disease Spread. *in*, *Spatial Components of Plant Disease Epidemics*, ed. M.J. Jeger. New Jersey: Prentice-Hall, Inc. pp. 127-143.

Mundt, C.C. (1989) Use of the modified Gregory model to describe primary disease gradients of wheat leaf rust produced from area sources of inoculum. *Phytopathology* 79(2):241-246.

Murdoch, W.W. (1975) Diversity, complexity, stability, and pest control. *Journal of Applied Ecology* 12:795-807.

Perrin, R.M. (1980) The role of environmental diversity in crop protection. *Protection Ecology* 2:77-144.

Reynolds, K.M., and L.V. Madden. (1988) Analysis of epidemics using spatio-temporal autocorrelation. *Phytopathology* 78:240-246.

Riitters, K.H.R. (1994) Landstat: landscape ecology research software.

Riitters, K.H.R., R.V. O'Neill, C.T. Hunsaker, J.D. Wickham, D.H. Yankee, S.P. Timmins, K.B. Jones, and B.L. Jackson. (1995) A factor analysis of landscape pattern and structure metrics. *Landscape Ecology* 10(1):23-39.

Risser, P.G., J.R. Karr, and R.T.T. Forman. (1984) *Landscape ecology: directions and approaches*. Illinois Natural Historic Survey Special Publication, No. 2.

Roberts, D.A., and C.W. Boothroyd. (1975) *Fundamentals of plant pathology*. San Francisco: W.H. Freeman and Company.

Turner, M.G. (1989) The effect of pattern on process. *Annual Review of Ecological Systems* 20:171-197.

Turner, M.G., ed. (1987) *Landscape heterogeneity and disturbance*. New York: Springer-Verlag.

Wegner, J., and G. Merriam. (1979) Movement by birds and small mammals between a wood and adjoining farm habitats. *Journal of Applied Ecology* 16:349-357.

James Newell
Institute of Applied Sciences
University of North Texas
Denton, TX 76203
jnewell@gab.unt.edu

Jeffrey Fitzgerald
Department of Geography
University of North Texas
Denton, TX 76203
fitz@unt.edu

T. Scott Rupp

LANDSCAPE-LEVEL MODELING OF SPRUCE SEEDFALL USING A GEOGRAPHIC INFORMATION SYSTEM.

ABSTRACT

This paper integrates two mathematical dispersal models into a GIS and outlines the initial development of a spatially explicit model for white spruce seed dispersal in interior Alaska. Integration of models revealed several benefits of modeling with a GIS. These benefits include simulation of dispersal upon a defined landscape unit, the ability to describe the effects of source shape upon dispersal patterns, simulation of the natural variation in seed density, and the application of dispersal simulations to aid resource managers with silvicultural decisions. Development of a spatially explicit model for interior Alaska demonstrates the potential for simulating the effects of seed crop periodicity and the use of a wind friction parameter to describe the influence of wind upon general dispersal patterns. The lack of data for interior Alaska is a major limitation to the modeling effort. This model is still under development, but at the very least demonstrates the potential of a GIS for environmental modeling of spatially explicit processes such as seed dispersal.

INTRODUCTION

Understanding the controls upon landscape-level vegetation patterns is crucial to successful ecosystem management. An in-depth knowledge of regeneration dynamics is a prerequisite to such an understanding. Seed dispersal patterns are a critical factor affecting colonization patterns. These patterns directly influence subsequent processes within the regeneration phase. Therefore, a spatially explicit understanding of dispersal will help our understanding of controls upon these landscape patterns.

The landscape of interior Alaska is characterized by a mosaic of forest types. The fire history of the last 200 years is closely related to this vegetation mosaic (Viereck 1973). Recolonization of a burned site depends on several factors including the availability of seeds from adjacent communities (Zasada et al. 1983). The ability to model dispersal patterns upon the landscape would improve our knowledge of boreal forest regeneration dynamics and provide land managers with a useful management tool.

Improved understanding of seed dispersal mechanics would be a great benefit to the land manager. Past studies have sought to mechanistically model the dispersal process by identifying parameters such as height of release, settling velocity, wind speed and turbulence, and seed structure (Greene and Johnson 1989 and 1995, Okubo and Levin 1989, and Sharpe and Fields 1982). The advantage of such models are the ability to apply the model to any species or situation. This approach differs from the traditional area specific physical parameters of numbers of seed and distance from the source which must be measured for each specific situation (Okubo and Levin 1989). Mechanistic models of point source dispersion have been developed and show promise, but area source models have yet to be developed at a practical level useful to the land manager.

Traditional modeling studies of seed dispersal for most tree species have been largely restricted to mathematical expressions of seed dispersed, as a function of distance from the source, into a spatially undefined clearcut opening. Studies involving seed traps in clearcuts have been conducted by researchers for decades. Dobbs (1976) and Youngblood and Max (1992) investigated seed dispersal of white spruce. These studies have provided valuable information on seed production, dispersal distances, and subsequent seedling densities. Although these studies describe seed dispersal they lack the ability to be applied to a spatially defined landscape.

Geographic information systems (GIS) offer the ability for simulating general seed dispersal patterns upon a spatially defined landscape. This approach allows for simulation of variability in seed numbers at a given distance, effects of wind upon seed distribution, and influences of seed source shape. Although such a model is not as powerful and robust as the above mentioned mechanistic approach, it does provide a silvicultural management tool of an applied nature for the resource manager.

Our goal is to model regeneration of interior Alaska white spruce forests, solely within a GIS. The regeneration routine is being developed for integration into a GIS based forest ecosystem model. This paper describes the initial subroutine development of seed dispersal. The development process began by integrating previous models into a GIS. This integration effort was utilized to investigate the potential advantages of modeling within a GIS. The actual subroutine development of white spruce seed dispersal in interior Alaska is ongoing. The dispersal routine will model seedcrop periodicity and the subsequent dispersal patterns from the seed source into a post-disturbance clearing.

MODEL INTEGRATION

The dispersal distance of seeds into a clearcut is an important factor in the ability of natural regeneration to provide adequate restocking levels. Informed decisions regarding clearcut size, seedbed preparation, and whether or not to regenerate artificially depend on having a good measure of seed dispersal densities (Dobbs 1976). Integrating dispersal equations of previous white spruce seedfall studies into a GIS can provide more detailed information for such decisions. Two different dispersal equations were used to investigate the benefits of modeling dispersal on a GIS. The model was developed as an ARC/INFO arc macro language (AML) routine within the GRID package.

Our initial code used Dobbs (1976) quadratic equation for white spruce seed dispersal in central British Columbia. The objective was to disperse seeds on a defined landscape unit and simulate seedfall differences associated with various source shapes.

Dispersal of seed upon a defined landscape provides specific information about a physical location. The AML inputs a grid identifying the source cells. The euclidean distance of each cell to the closest source cell is then determined. The distance grid is then applied to the dispersal equation from which an output grid defines seed numbers for each cell. Describing the source shape allows for simulation of dispersal patterns specific to the seed source (Figure 1). This model provides information about seed density upon a defined landscape unit and the effects of source shape which the original equation of Dobbs (1976) cannot provide.

Dispersal of interior Alaska white spruce seed was also integrated. The negative exponential equation, of Youngblood and Max (1992), for floodplain white spruce seed dispersal in interior Alaska was utilized. The objectives were to simulate seed density variance at a given distance and to provide information which can be utilized by the resource manager to assist in silvicultural decisions.

The ability to describe seed density variance upon the landscape provides another layer of information. Using the original data set, the variance in seed numbers as a function of distance was calculated. Within a given distance seed was dispersed upon the cells by applying a normal distribution with the model mean and standard deviation associated with that distance interval (Figure 2).

The above information, coupled with information on seed to germinant ratios and stocking requirements, would allow the manger to obtain a visual model of potential natural regeneration following a harvest of given size and shape. Using information from Zasada (1971) a 24:1 seed to germinant ratio was applied to create a grid of successful germinants for each cell. Full stocking was assumed to represent 2500 trees/ha (Cleary et al. 1978). A stocking grid was then created to identify four stocking levels as defined by Reynolds et al. (1953). Stocking regions are developed through the use of several focal functions. Analyzing neighboring cells by calculating mean values and using majority grouping allows for a description of generalized stocking patterns. The output grids allow for the analysis of potential stocking following a harvest (Figure 3). This model provides a more detailed and realistic description of dispersal upon the landscape than the original model of Youngblood and Max (1992) and provides information that can be applied by the resource manager in silvicultural decisions.

MODEL DEVELOPMENT

Integration of seed dispersal equations into a GIS allows for modeling on a spatially defined landscape. The dispersal routine simulates both the periodic nature of seed production and the subsequent dispersal of the seedcrop. The work of Youngblood and Max (1992) is being incorporated with current field study work to model seedfall patterns as a function of seedcrop quality, distance from the source stand, and wind direction.

White spruce cone crops are periodic in nature, being influenced heavily by annual temperature and precipitation regimes (Zasada 1971). Cone crop periodicity is modeled with a Monte Carlo simulation technique following the work of Fox et al. (1984). The subroutine simulates the probabilistic nature of white spruce cone crops. The seed crop is classified as either good-excellent or poor-moderate. A probability of 0.25 for a good-excellent seed year was projected (Zasada and Viereck 1970 and Zasada 1980). Furthermore, successive good-excellent seed years do not occur and the interval between them is irregular (Zasada 1971, 1980). This periodic nature has a big impact on potential stocking levels. Seed densities associated with cone crops were estimated based on limited data available (Zasada and Viereck 1970). The AML subroutine used scalars and the random number function to simulate seed crop quality. Seed density is then calculated from the best estimated mean and standard deviation for that cone crop rating. This variable is then input into the dispersal AML. By determining seed density ranges associated with the two quality classes, cone crop influences on dispersal density and associated stocking levels can be described (Figure 4).

Winds influence general dispersal patterns. Most studies have implicitly modeled this influence by deploying seed traps from the windward edge into the clearing in the direction of the prevailing winds. The ability to model wind influences in a simplistic manner would benefit the resource manager. This model attempts to use a wind friction parameter to describe the negative influence upon general dispersal in non- windward directions. The parameter reduces the distance and associated density of seeds dispersing in directions from the source other than that of the defined wind direction. This parameter allows for the potential simulation of wind patterns upon general seed dispersal patterns. The development of a wind friction parameter is ongoing and much of the field work portion has yet to be completed.

This first generation dispersal subroutine for interior Alaska white spruce is still in a developmental stage. Due to the relative lack of data much field work must be completed before the dispersal routine can be finished and tested.

DISCUSSION

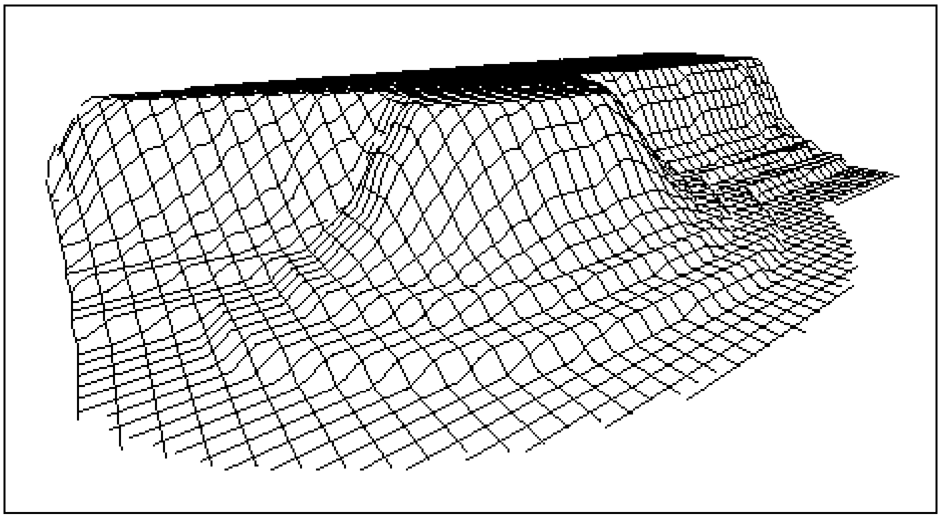
The ability to apply a model upon a defined landscape unit is an improvement over most modeling efforts which work within an undefined landscape. Current GIS packages, such as ARC/INFO, have many functions and utilities that lend itself to developing simple AML's that can model certain biological processes upon a given landscape. This paper provides such an example. Integration of prior seed dispersal models demonstrates the advantages of modeling in a GIS and the model development shows the potential for development of a silvicultural tool easily applied by the resource manager to aid in decision making.

Integration of mathematical models of white spruce seed dispersal into a GIS was relatively easy. The modeling effort provided information about the dispersal process beyond that of the original mathematical models. The most identifiable advantage is the ability to disperse seeds upon a specific landscape unit and provide information about a physical location. This point is further demonstrated by the ability to describe source shape and its influence upon dispersal patterns. Providing dispersal distributions associated with seed number deviations at a given distance interval relates a more realistic view of the dispersal process. Utilizing all this information, along with stocking level information, demonstrates the potential for use as a silvicultural tool by the resource manager. Information on disturbance size and shape and possible stocking levels could help in the development of specific harvesting schemes and decisions.

The seed dispersal routine for interior Alaska incorporates seed crop periodicity, providing description of another important factor in the seedfall dynamics of white spruce. Modeling the influence of winds upon dispersal patterns will further benefit the resource manager.

Application of a wind friction parameter can help describe seedfall patterns and help in the creation of harvesting schemes and making silvicultural decisions about stocking levels. The lack of seedfall data sets for Alaska white spruce is a major limitation. Data acquisition will provide for further model development and testing.

This paper provides an idea of the potential of a GIS for modeling biological processes upon the landscape and for use of such models by the resource manager to aid in decision making. Benefits of modeling with a GIS include the ability to simulate biological processes upon a defined landscape unit, provide for natural variation across the landscape, and provide simulations for decision making purposes.



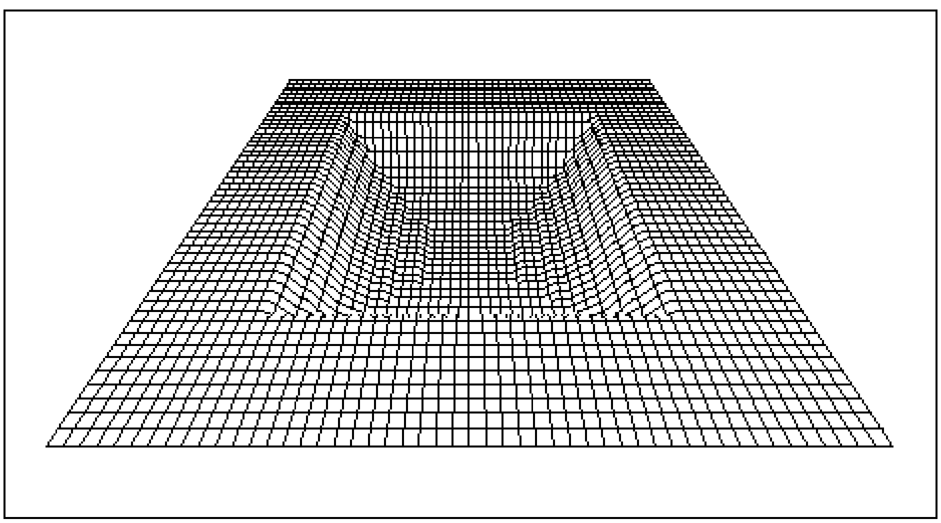


Figure 1 -

Influence of source geometry upon dispersal patterns. Seed density is displayed as a surface upon the landscape. Areas of greatest elevation identify the source. Dispersal AML uses equation from Dobbs (1976).

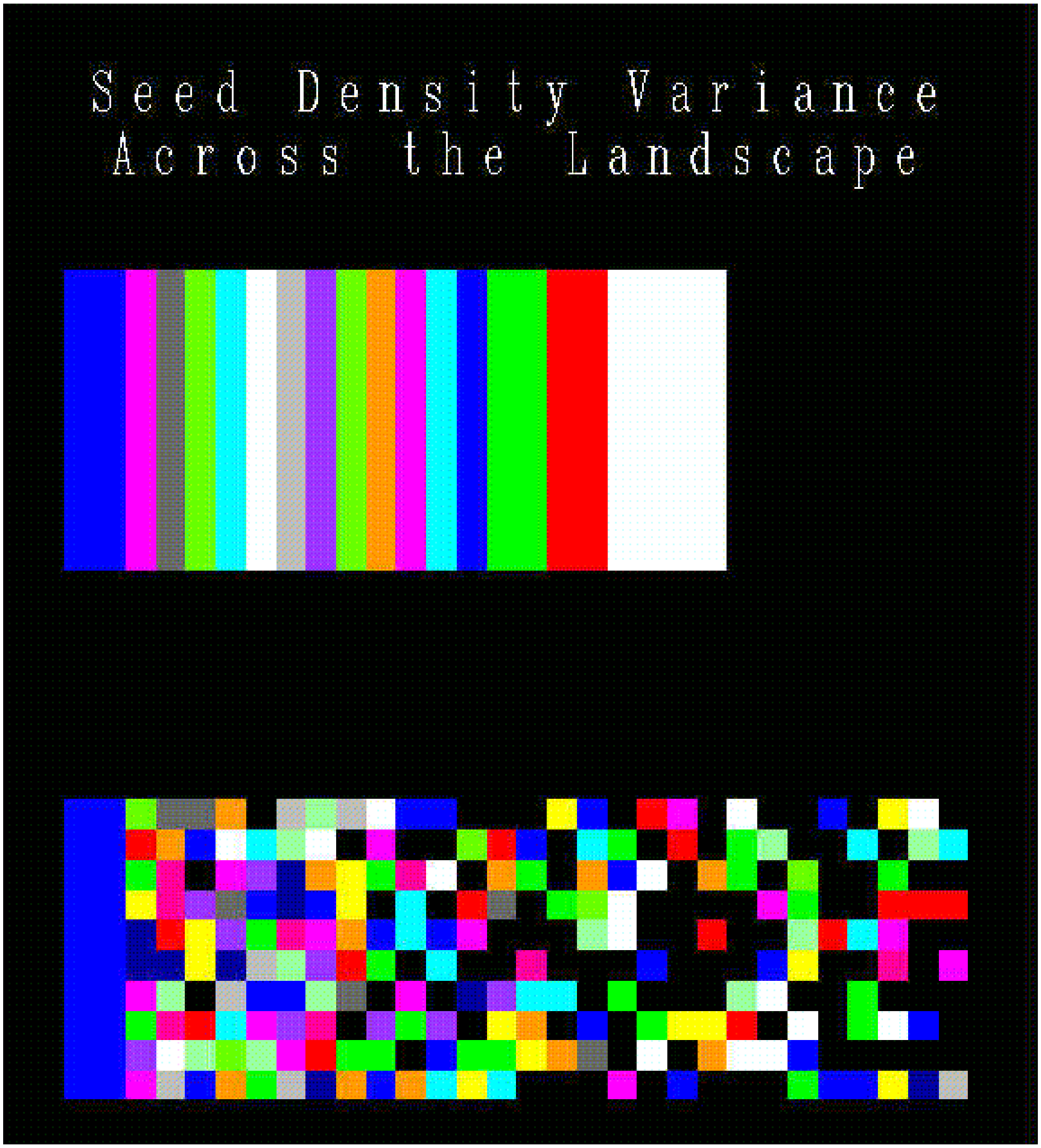


Figure 2 -

Output grids of dispersal AML using the equation of Youngblood and Max (1992). Colors represent the percent values of filled seed dispersed at a given distance from that dispersed within the source. A) Output grid from negative exponential model. B) Output grid showing density variance within a given distance interval.

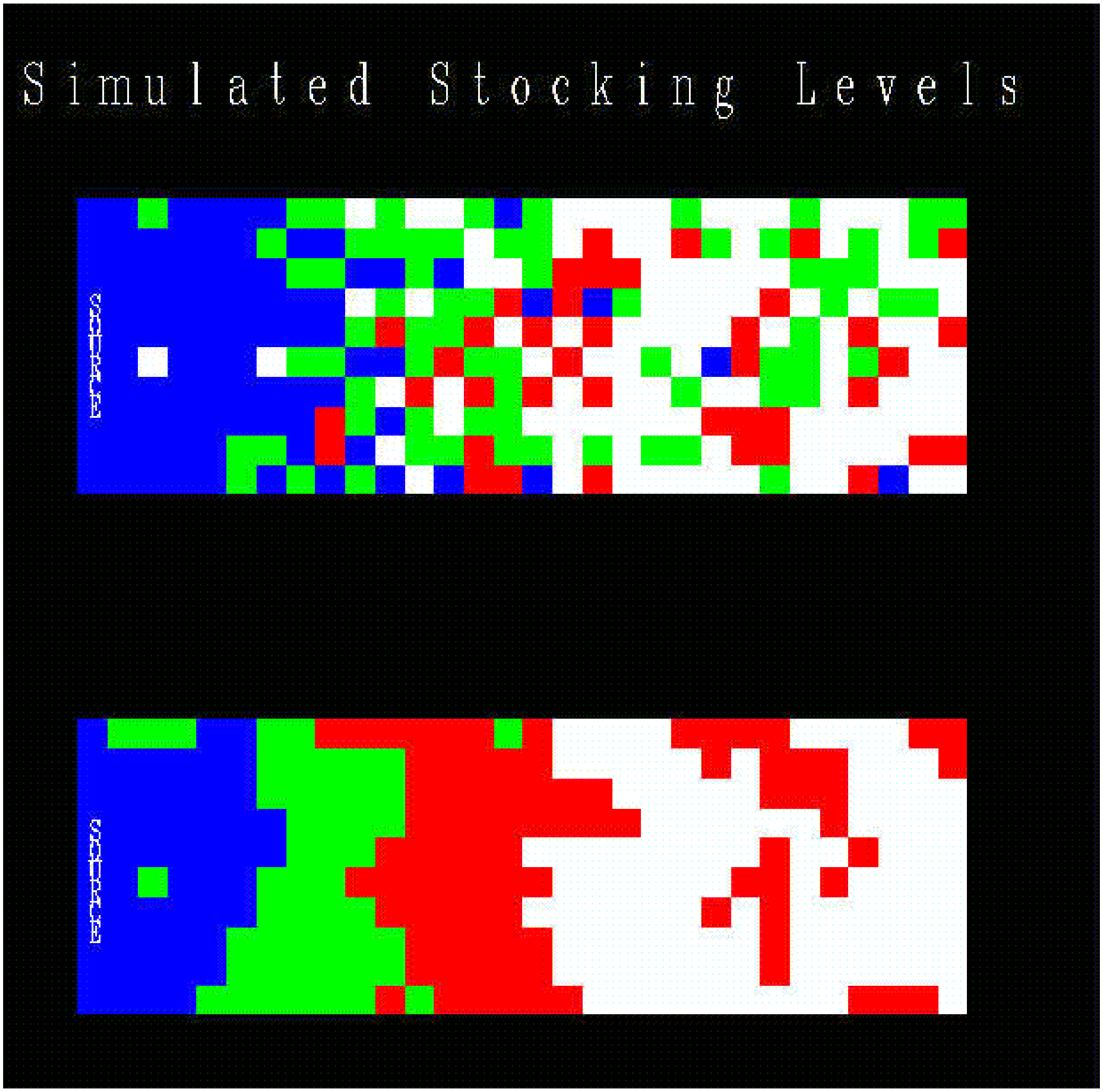


Figure 3 -

Output grid cell color values represent stocking levels (Blue = good, Green = medium, Red = poor, White = nonstocked). A) Initial stocking grid calculated from successful germinants. B) General stocking regions as determined by focal functions application.

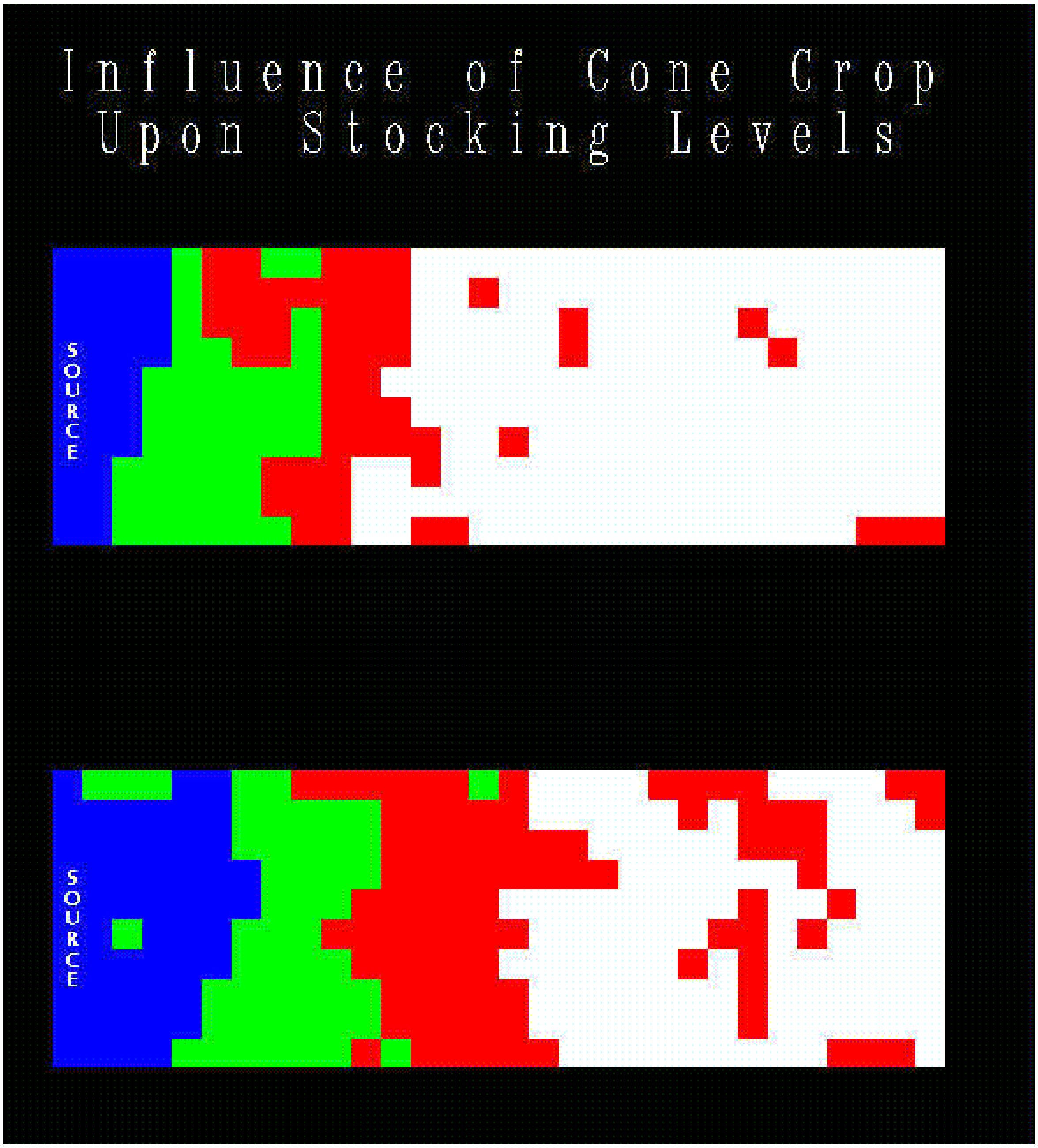


Figure 4 -

Stocking regions for different seed crop classifications. Simulates the effects of cone crop periodicity upon regeneration. A) General stocking regions resulting from a poor-moderate cone crop. B) Stocking regions resulting from a good-excellent cone crop.

REFERENCES

Cleary, B.D., Greaves, R.D., and Hermann, R.K. (1978) Regenerating Oregon's forests.

Oregon State Univ., Ext. Serv., Corvallis, OR.

Dobbs, R.C. (1976) White spruce seed dispersal in central British Colombia. *For. Chron.* 52: 225-228.

Fox, J.D., Zasada, J.C., Gasbarro, A.F., and Van Veldhuizen, R. (1984) Monte Carlo simulation of white spruce regeneration after logging in interior Alaska. *Can. J. For. Res.* 14: 617-622.

Greene, D.F., and Johnson, E.A. (1989) A model of wind dispersal of winged or plumed seeds. *Ecology*. 70: 339-347.

Greene, D.F., and Johnson, E.A. (1995) Long-distance wind dispersal of tree seeds. *Can. J. Bot.* 73: 1036-1045.

Okubo, A., and Levin, S.A. (1989) A theoretical framework for data analysis of wind dispersal of seeds and pollen. *Ecology.* 70: 329-338.

Reynolds, C., Jeffers, N., Borisquet, V., and Stien, R. (195 3) Reforestation surveys. Chapter VII. In *Reports of the Pacific Northwest seeding and planting committee on various recommended reforestation practices and techniques.* pp. 61- 69.

Sharpe, D.M., and Fields, D.E. (1982) Integrating the effects of climate and seed fall velocities on seed dispersal by wind: a model and application. *Ecol. Mono.* 17: 297-310.

Youngblood, A., and Max, T.A. (1992) Dispersal of white spruce seed on Willow Island in interior Alaska. *Res. Pap. PNW-RP-443. Portland, OR. USDA For. Serv., Pacific Northwest Res. Sta.*

Viereck, L.A. (1973) Wildfire in the taiga of Alaska. *Quatern. Res.* 3: 465-495.

Zasada, J.C., and Viereck, L.A. (1970) White spruce cone and seed production in interior Alaska, 1957-1968. *USDA For. Serv. Res. Note PNW-129, Pac. Northwest For. Range Exp. Stn., Portland, OR.*

Zasada, J.C. (1971) Natural regeneration of interior Alaska forests - seed, seedbed, and vegetative reproduction considerations. In *Fire in the northern environment, a Symposium. Pac. Northwest For. Range Exp. Stn., Portland, OR.* pp. 231-246.

Zasada, J.C. (1980) Some considerations in the natural regeneration of white spruce in interior Alaska. In *Forest regeneration at high latitudes. Proceedings of an International Workshop. USDA For. Serv. Gen. Tech. Rep. PNW-107, Pac. Northwest For. Range Exp. Stn., Portland, OR.* pp. 25-29.

Zasada, J.C., Norum, R.A., Van Veldhuizen, R.M., and Teutsch, C.E. (1983) Artificial regeneration of trees and tall shrubs in experimentally burned upland black spruce/feather moss stands in Alaska. *Can. J. For. Res.* 13: 903-913.

T. Scott Rupp

Doctoral Research Assistant
Forest Soils Laboratory
University of Alaska
Fairbanks, AK 99775
Telephone: (907) 474-7019
FAX: (907) 474-6184
srupp@salrm.alaska.edu

The Effects of Elevation Data Representation on Mesoscale Atmospheric Model Simulations

Hoyt Walker, John M. Leone, Jr. and Jinwon Kim
Lawrence Livermore National Laboratory
P.O. Box 808
Livermore, CA 94551-0808
(510) 422-1840
Fax: (510) 423-4527

ABSTRACT

Mesoscale atmospheric model simulations rely on descriptions of the land surface characteristics, which must be developed from geographic databases. Certain features of the geographic data, such as its resolution and accuracy, as well as the method of processing for use in the model, can be very important in producing accurate model simulations. The work described here is part of a research effort into the relationship between these aspects of geographic data and the performance of mesoscale atmospheric models and is particularly focused on elevation data and how it is prepared for use in such models.

A source for digital elevation data will typically not be at the resolution required for a given model simulation and so a resampling step is required. In addition, predictive non-linear model often cannot accept forcing at high spatial frequencies due to the terrain, thus smoothing is also required. The effect of different means of resampling and smoothing elevation data on two types of model simulations is investigated. At smaller spatial scales, nocturnal drainage winds in mountain valleys in Colorado are examined for effects on the general characteristics as well as the details of the flows. At the larger end of the mesoscale, extended simulations of California weather are examined for effects on orographic lifting, low-level convergence and divergence and ultimately rain and snow distribution. In both of the situations, the terrain representation can have significant effects on the simulated flow that could be important in some applications.

INTRODUCTION

In mesoscale atmospheric modeling, a variety of surface features such as the elevation, surface roughness, and sensible and latent heat fluxes must be represented in the model (Pielke, 1984; Lee, et.al., 1991). These features, expressed in the form of model boundary conditions, must be developed from geographical databases in a way that balances the need for descriptive detail and accuracy with the spatial discretization of a specific model simulation. This is due to both the physics and numerical techniques in a model and to the close coupling of the land surface with boundary layer phenomena under many conditions (Pielke, 1984). This paper focuses specifically on elevation data and how it is processed for use in an atmospheric model. Because of its static nature, it forms a convenient starting point for looking at the interaction between geographic data and atmospheric models. Before describing the details of the research presented in this paper, it is appropriate to provide some background on the use of elevation data in regional atmospheric models.

Elevation data representation in mesoscale atmospheric models.

Defining the shape of the lower boundary of an atmospheric model domain using elevation data is clearly an essential component for explicit modeling of the atmospheric boundary layer given the need to represent the steering and lifting effects of the elevation surface, as well as to model the variable surface heating due spatial variations in terrain slope and aspect. There are a variety of ways of defining the elevation surface in a model, which for mesoscale models involves defining the elevations at the horizontal grids points of the model (global models often solve their governing equations in spectral space and consequently descretize spectral space, in contrast, mesoscale models are essentially always expressed in physical space with model variables tracked on a three-dimensional grid). For example, triangular or rectangular grids can be used. Although research into triangular grids is on-going (see Boybeyi, et.al., 1994), virtually all mesoscale models in common use rely on rectangular grids. Terrain representations can also be classified into stepped and continuous representations. In stepped terrain, the landform is approximated by stacks of building blocks or tiles that represent some average elevation over a model grid cell, while in a continuous representation, the elevations are defined at horizontal grid points and some interpolation scheme, e.g., bilinear, can be assumed to define a continuous surface at all x,y locations in the model grid. These approaches are illustrated in Figure 1 for two dimensions. The National Weather Service Eta model uses a stepped terrain representation operationally to perform weather prediction for North America (Mesinger, et.al., 1988). As most mesoscale atmospheric models use some form of continuous terrain representation, this paper concentrates on models that use rectangular grids and continuous terrain representations.

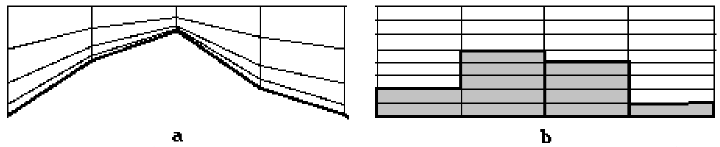


Figure 1. Two-dimensional examples of (a) a continuous terrain representation with terrain following coordinates and a variable vertical grid step and (b) a stepped terrain representation with a constant vertical grid step.

Elevation data has a particularly critical role in most mesoscale models, not only because it defines the shape of the lower boundary but the landform also affects the spatial discretization throughout much of the grid volume as shown in Figure 1a. This is because such models typically incorporate some form of terrain-following coordinate system where the terrain surface is defined as one limit of a transformed vertical coordinate (e.g., 0) and the top of the model domain is defined as the other limit (e.g., 1). In physical coordinates, the grid points are compressed and stretched as the terrain rises and falls, again as shown in Figure 1a. Thus the landform not only affects the details of the surface interactions, but it also affects the geometry of the grid above the surface. Therefore, the manner in which the terrain is represented can have effects on the numerical representation of the flow throughout much of the grid. Also, shown in Figure 1a is a graded (or stretched) vertical grid, where there is a greater density of grid points near the land surface and fewer points near the grid top. Given the complexity of the physical processes just above the land surface and their relatively small length scale, using greater resolution near the ground is usually necessary in mesoscale models that explicitly model the behavior of the planetary boundary layer.

An important constraint on prognostic model representations of terrain is related to stable performance through simulated time of such models. Forcing a non-linear prognostic model at high spatial frequencies relative to the model grid step (i.e. variation with wavelengths of 2-4Δx, where Δx is the length of a model grid step) can destabilize model performance, which can result in the exponential buildup of high-frequency waves that swamp the realistic aspects of the simulation. Winds blowing over terrain with significant 2-4Δx variation represents a constant high frequency forcing of the model that can limit the accuracy of the simulation. To avoid this problem, the model elevations are typically smoothed. A related issue involves the parameterization of subgrid scale terrain variation. Any roughness on the earth's surface acts as a momentum sink and thus affects the atmospheric flow. The elevation surface in the model can only represent and explicitly model the effects of terrain variation with wavelengths of 2Δx and longer. Unresolved terrain variation is can be parameterized by adjusting the surface roughness height, which is normally used to express the effect on the flow of surface elements such as grass, trees and buildings (see, for example, Georgelin, et.al., 1994; Mason, 1991). Another important aspect of terrain representation is associated with the barrier that a range of hills or mountains presents to an atmospheric flow across it. A number of important atmospheric phenomena are strongly affected by the height of such a barrier, e.g., lee cyclogenesis and orographic precipitation. Problems can occur with a terrain representation that relies on a simple mean to represent the barrier height if the model grid step is large with respect to the horizontal scale of significant relief. Mean elevations tend to lower the maximum heights, thus altering the effective barrier height for the flow. At synoptic and larger scales, mean elevations are often modified by adding a factor (typically 1 or 2) times the standard deviation of the terrain. Such a surface is called an envelope terrain and it raises the effective barrier and produces better model simulations (Tibaldi, 1986; Wallace, et.al., 1983). These examples illustrate the need to consider the physics, the numerics and the application in determining how to most effectively representation geographic data in an atmospheric model.

This research focuses on two aspects of elevation data representation in mesoscale atmospheric models. First, the sensitivity of small-scale model simulations of nocturnal drainage winds to the smoothness and variability of the elevation data representation is examined The development of nocturnal drainage winds is one example of a terrain driven atmospheric flow and offers a useful test case for examining the effects of elevation data on mesoscale models (Leone & Lee, 1989). In earlier work, a number of model simulations using idealized terrain have been performed and analyzed (Walker & Leone, 1994). Here we extend the investigation by conducting a series of model experiments to determine the response of a hydrostatic mesoscale atmospheric model to lower boundary forcing due to variations in the representation of a real mountain valley system during nocturnal cooling.

Second, the effect of different resampling techniques on a regional-scale simulation of precipitation is considered. There are numerous ways of resampling elevation data for regional models that are intended to resolve primarily sub-synoptic scale motions. Perhaps the most common method is to use a grid cell mean terrain computed from fine resolution elevation data. As suggested by the synoptic scale work mentioned above, the orographic effects of a mountain barrier can sometimes be improved using an envelope terrain, which can differ substantially from the mean terrain in mountainous regions. For mesoscale motions, the terrain affects the atmospheric flow through orographically-generated vertical motion and local convergence, which ultimately affect the low-level wind and the transport of tracers, such as water vapor, by the low-level wind. The effects of using these two approaches to creating a terrain representation will be examined here.

For historical reasons, these two aspects of this study have been performed using different mesoscale models. As a result, the rest of the paper will be structured into sections covering the small-scale effects, followed by a section covering the regional-scale effects. The results of both sections will be summarized in the conclusion.

SMALL-SCALE EFFECTS

In this study, a basic drainage flow is defined along the Brush Creek valley system in western Colorado, the site of the ASCOT field experiments (see Clements et.al., 1989), using a mesoscale atmospheric model relying on unsmoothed elevation data to define the lower boundary (for this particular situation, unsmoothed elevation data does not cause problems with the stability of the simulation). The elevation data is then altered in various ways and used as the basis for additional simulations. Some specific details of the control simulation and one comparison run are described in the following sections.

Model Description

The atmospheric model used in these tests is called SABLE (Simulator of the Atmospheric Boundary Layer Environment), a hydrostatic mesoscale model developed at the Lawrence Livermore National Laboratory. SABLE solves the hydrostatic, anelastic, equations for velocity, potential temperature, and Exner function in three dimensions (Zhong, et.al. 1991). The equations are solved by using a unique blend of numerical techniques. The prognostic equations for the horizontal velocity components and the potential temperature are solved using trilinear, isoparametric finite elements in space combined with a semi-implicit time integration scheme. The diagnostic equations for vertical velocity and Exner function are solved by integrating up or down vertical columns, respectively, via centered finite differences. Turbulence was modeled using the local Richardson number-dependent K model of McNider and Pielke (1981).

Model Domain

For all simulations, a horizontal grid was used that covers a 7 by 32 km area oriented along Brush Creek with a 200 m grid step, , i.e., the grid had 36 by 161 nodes in the horizontal directions (see Figure 2). The domain also includes a portion of the valley into which Brush Creek drains (Roan Creek). The horizontal coordinate system was derived from the Universal Transverse Mercator (UTM) projection for this area using a translation and 45 degree rotation to align the y axis with the centerline of Brush Creek. The upper boundary of the domain was is flat at an altitude of 4000 m (the minimum and maximum elevations in the grid for the unsmoothed terrain are 1650 and 2650 m, respectively). The vertical grid was graded with the lowest cell being 20 m.

The lower boundary of the domain was defined using elevation data extracted from a Defense Mapping Agency (DMA) Digital Terrain Elevation Data (DTED) quadrangle covering the Brush Creek area at 3 arc-second resolution. The raw elevation data was resampled to the model grid using an unweighted mean. The smoothed terrain was generated by using a 9 by 9 binomial filter data with only the original data used at each step in computing the weighted average to avoid propagation effects (see Figure 2b). Elevations beyond the model grid boundary were accessed to create a buffer regional so that the filter stencil could extend beyond the model grid without using an artificial boundary condition. Sine waves were added to the valley floor and sides for two of the runs with the waves diminishing to zero as the valley ridges were approached. The magnitudes of the sine waves were 40 m and two wavelengths were used, $4\Delta x$ and $8\Delta x$ (see Figures 2c and 2d).

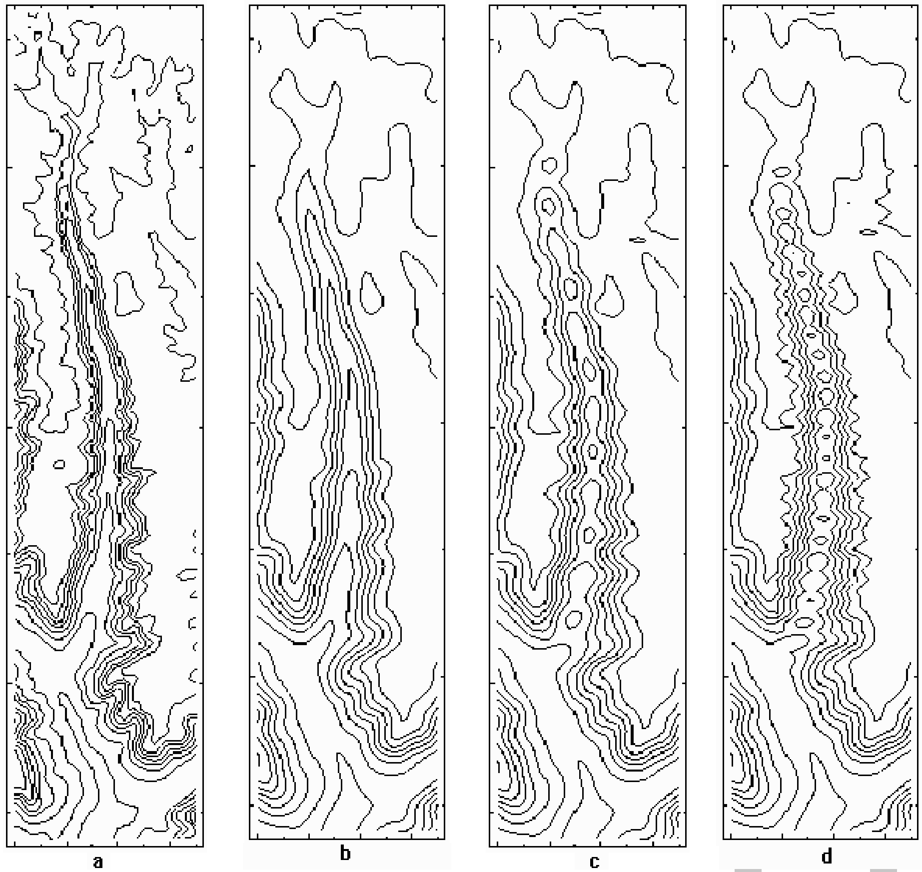


Figure 2. Contours of the elevation data use for the (a) unsmoothed, (b) smoothed, (c) 8 Δx and (d) 4 Δx simulations. The contour interval is 100 m with the lowest contour at 1700 m.

Initialization

Given the goal of isolating the effects of terrain representation on a model run, accurately reproducing any particular physical situation was not of great importance. Thus, a number of simplifying assumptions were made. For example, the Coriolis parameter was set to zero to avoid complicated veering motions. The cross-side valley wind component, u, was assumed to be zero at the appropriate lateral boundaries. At the top boundary, both horizontal wind components, u and v, were set to zero. The lower boundary cooling was specified as a heat flux of -60 W/(m*m). The atmosphere was initialized to be slightly stable with a potential temperature lapse rate of 0.002 K/m. The problems were run for 8 hours. These values were used in all of the runs, thus, the only differences between simulations were the elevation surfaces.

Results

The unsmoothed terrain was successfully integrated for 8 hours. The cooling land surface caused the generation of downslope winds that develop a distinct jet that gained strength as it moved down Brush Creek until it approached the intersection with the Roan Creek valley. Here, the jet leveled off and began to diminish as it neared the end of the grid. Wind vectors at grid points three levels above the surface are shown in Figure 3a, which illustrates the flow at 4 hours into the simulation. Figure 3b provides the same information, but at 8 hours, and shows a similar flow pattern over most of the length of the valley; however there is a noticeable lessening of the flow at the mouth of the valley and into the Roan Creek Valley. This is related to the pooling of cold air in the lowest areas of the valley system. The variation of the maximum down-valley wind component along the valley is summarized in Figure 3c.

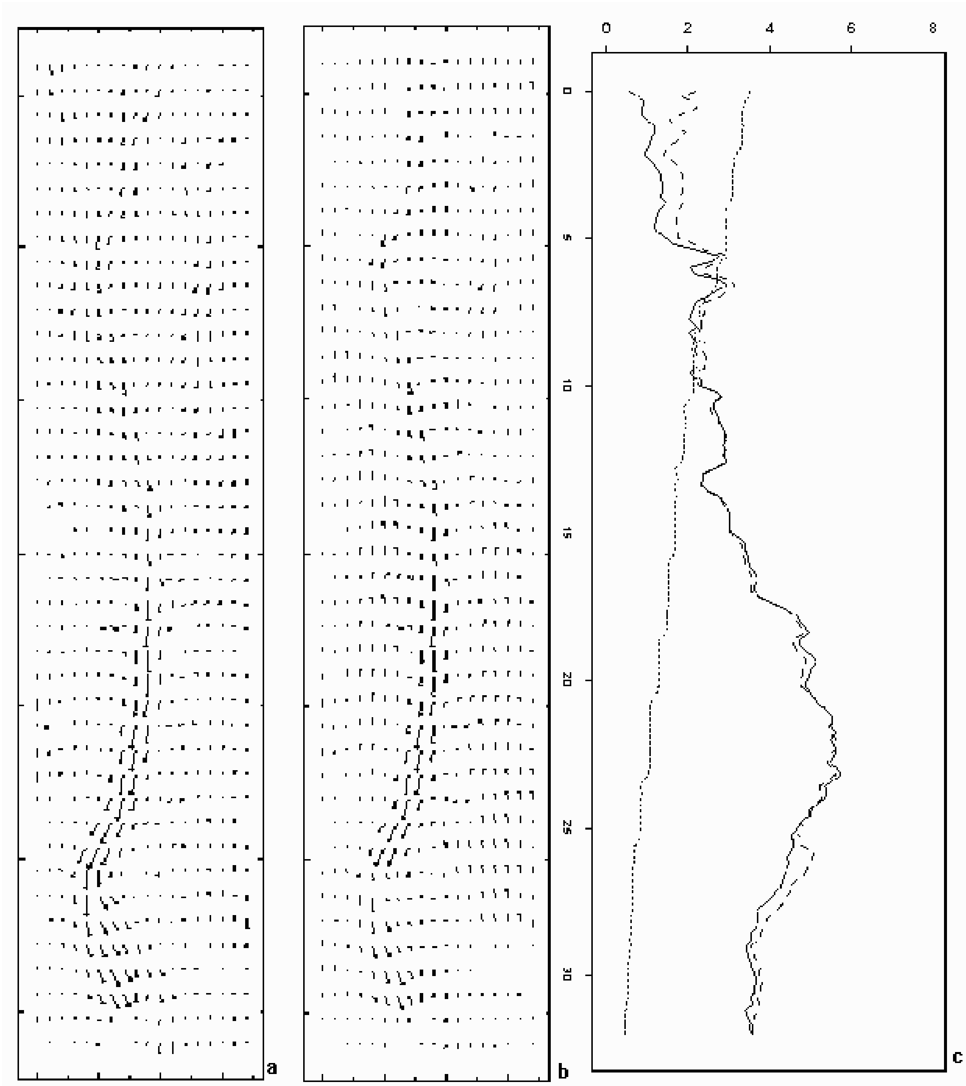


Figure 3. Wind vectors three grid levels above the terrain for the control run at (a) 4 hours and (b) 8 hours. Every other point is plotted along the x-axis and every fourth point is plotted along the y-axis. Maximum down-valley wind speeds as a function of down-valley distance are plotted in (c). The dashed line is the maximum jet speed at 4 hours and the solid line is the speed at 8 hours. The dotted line is the elevation of the valley bottom.

The smoothed terrain simulation shows the same general characteristics as the unsmoothed terrain; however, the flow over the smoothed terrain is distinctly stronger. For example, the maximum speed along the length of the valley at 4 hours increases from 5.8 to 7.2 m/s for the unsmoothed and smoothed terrains, respectively. At 8 hours the corresponding increase is from 5.6 to 6.8 m/s. This is also apparent when the down-valley speed maximum for the smoothed and unsmoothed terrains are compared (see Figures 4a and 3c). Of particular interest is the change in the counterflow that occurs between these two runs. In addition to the main jet that forms in the valley, a counterflow also develops above the valley over the entire width of the domain for all the simulations. This counterflow changes significantly with the altered surface representations (see Figure 5) with the unsmoothed terrain having the strongest counterflow. The addition of waves to the smoothed terrain weakens the counterflow relative to the smoothed terrain as well as effecting the details of the jet as it flows over the waves. As the counterflow builds up between 4 and 8 hours into the simulation it tends to confine the jet within the valley walls.

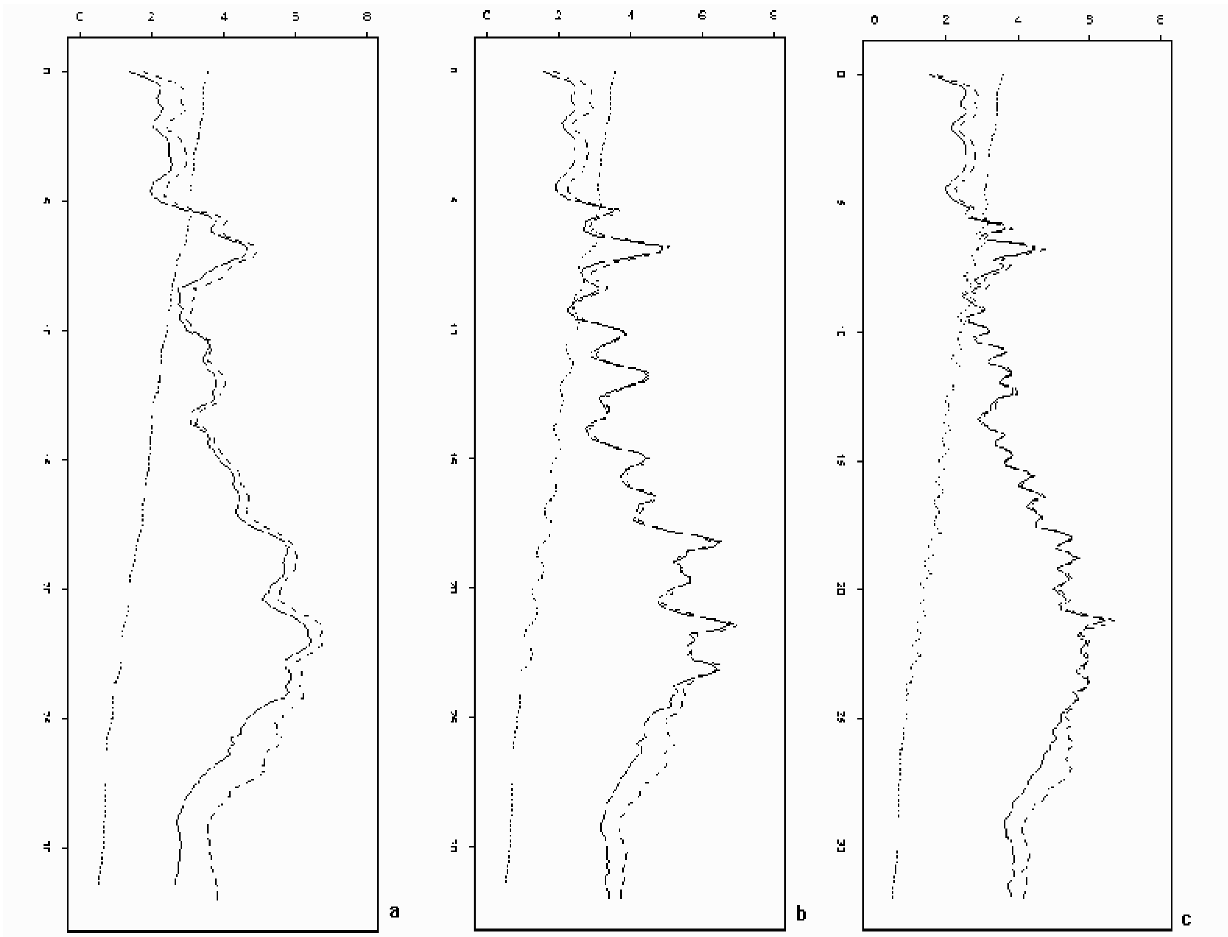


Figure 4. Maximum down-valley wind speeds as a function of down-valley distance for the (a) smoothed, (b) 8 Δx and (c) 4 Δx simulations. The dashed line is the maximum jet speed at 4 hours and the solid line is the speed at 8 hours. The dotted line is the elevation of the valley bottom.

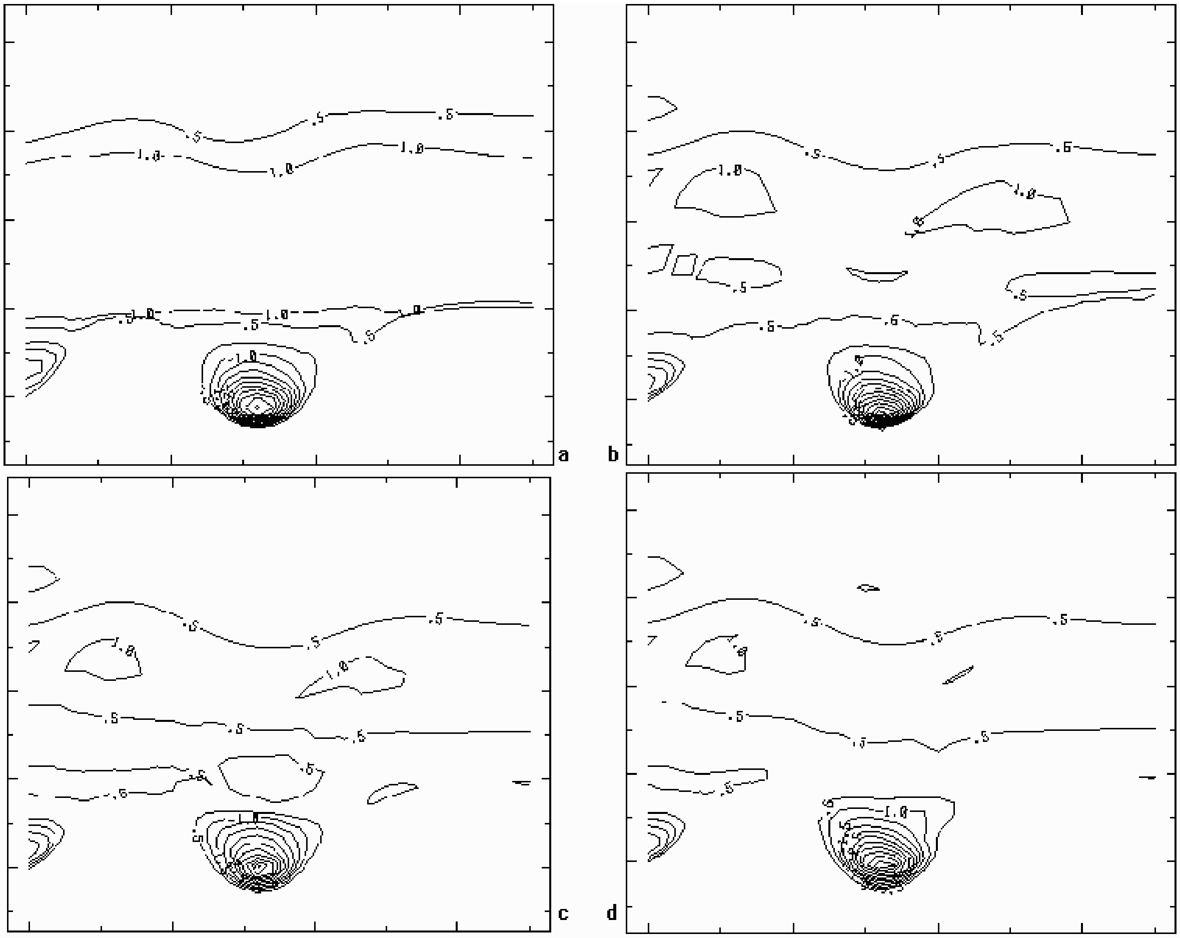


Figure 5. Contours of the down-valley wind component just before the maximum speed for the (a) unsmoothed, (b) smoothed, (c) $8\Delta x$ and (d) $4\Delta x$ simulations.

REGIONAL-SCALE EFFECTS

In this study, a winter storm in California is simulated using a primitive-equation, limited-area model with particular attention being focused on the distribution and intensity of precipitation in this area. The effects of two different terrain representations on the simulated low-level wind and precipitation and examined. The commonly used mean terrain representation provides the control case and a 1-sigma envelope provides a comparison simulation.

Model description

Details of the dynamical and physical formulations of the Mesoscale Atmospheric Simulation (MAS) model have been presented by Kim and Soong (1994) and Soong and Kim (1995). In summary, the governing equations of the MAS model are the flux-form of the primitive equations written in s-coordinates and discretized on Arakawa c-grid. The advection of momentum is computed using a third-order accurate scheme by Takacks (1985) with the advection of the remaining dependent variables computed by a finite difference scheme by Hsu and Arakawa (1990). Vertical staggering and differencing of variables follow the formulation by Arakawa and Suarez (1983).

Precipitation processes are computed separately for deep convection and grid-scale condensation using the Anthes cumulus parameterization and a bulk cloud microphysics scheme by Cho, et al. (1989), respectively, with these schemes integrated so as to conserve water and thermodynamic energy. Solar and terrestrial radiative transfer processes are calculated using multi-layer schemes (Harshvardhan et al. 1987). Effects of clouds on radiative transfer are computed separately for water- and ice-phase cloud particles using the formulations of Stephens (1978) and Starr and Cox (1982), respectively. Surface turbulent fluxes of momentum, heat, and water vapor are computed using the bulk aerodynamic transfer scheme (Deardorff, 1978). Vertical turbulent exchanges above the surface layer are computed using the K-theory. Drag coefficients at the surface and eddy diffusivities above the surface are computed using the formulation by Louis et al. (1981). Land surface processes are computed using the Coupled-Atmosphere-Plant-Snow (CAPS) model (Mahrt and Pan 1984; Kim et al. 1994; Kim and Ek 1995) that predicts soil water content and soil temperature and diagnoses the temperature and water vapor mixing ratio at land surfaces.

Model Domain

The computational domain covers a 1140 km x 1260 km wide region that contains the states of California, Nevada, and southern Oregon on the polar stereographic projection used by the NMC Eta model (see Figure 6). This area is covered with a 20 km x 20 km grid mesh in the horizontal. Fourteen irregularly-spaced layers between the ground surface and the top of the computational domain at the 50 mb level. The top of the computational domain was determined according to the availability of the NMC global analysis data to avoid extrapolating variables in the upper atmosphere. Enhanced horizontal and vertical diffusion was employed within the top three layers to reduce wave reflections at the rigid upper boundary. Additional five model layers are introduced between 50 mb and 1 mb levels to compute radiative transfer above the main computational domain.

Two terrain representations were used for the simulations. One was obtained by averaging fine-resolution (500 m) elevation data, derived from the DTED data, over each horizontal grid cell. This is referred to as the mean terrain. The second was obtained by adding the standard deviation of the elevations to the mean terrain for each horizontal grid cell with an additional check to ensure that the envelope terrain value was no greater than the maximum value within a horizontal grid cell. This is referred to as the envelope terrain. The envelop terrain enhanced the elevations along the Coastal Range and the Sierra Nevadas by 10-20%.

Initialization

The atmospheric variables were initialized by interpolating the 80 km resolution NMC ETA model initial data at 00UTC March 9, 1995 using the Cressman objective analysis scheme (Cressman 1959). Time-dependent lateral boundary conditions during the next two days were obtained by linearly interpolating over time between the NMC ETA-model initial fields available at 12-hour intervals.

Results

The low-level wind field at 18UTC March 9, 1995 (Figure 6) clearly illustrates the barrier effects of terrain on the low-level wind. The inflow from the Pacific Ocean turns to be increasingly parallel to the Coastal Range and the Sierra Nevada as it approaches these mountain ranges. The barrier jet, which is defined as the component of the wind parallel to the mountain (V in Figure 6) carries significant amount water vapor toward northern California. Hence, the strength of this barrier jet plays an important role in the precipitation that region.

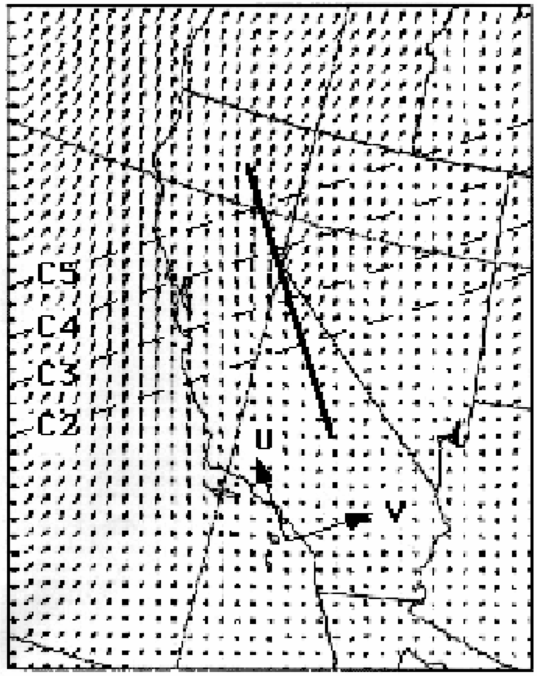


Figure 6. Wind vectors of low level winds for the California simulation at 1800UTC on March 9, 1995 also showing the location of the cross-sections used in Figures 7 and 8.

Figures 7 and 8 compare the v-wind component at four cross-sections (C2, C3, C4, and C5 in Figure 6). At all four cross-sections, the envelope terrain enhanced the low-level jet, appearing a short distance west of the peak of the Sierra Nevadas, by more than 2.5 ms-1 while the effects on the upper-level wind was small. Consequently, the low-level moisture transport into northern California region is enhanced with the envelope terrain. The envelope terrain also enhances the vertical motion along these cross-sections by about 10-30% (not shown). The enhanced low-level moisture transport and vertical motion appear to have increased precipitation in northern California (Figure 9). The most significant enhancement of local precipitation occurred at northern Coastal Range and northern Sierra Nevadas. Enhanced terrain along the southern Coastal Range also significantly increased local precipitation south of the Monterey Bay.

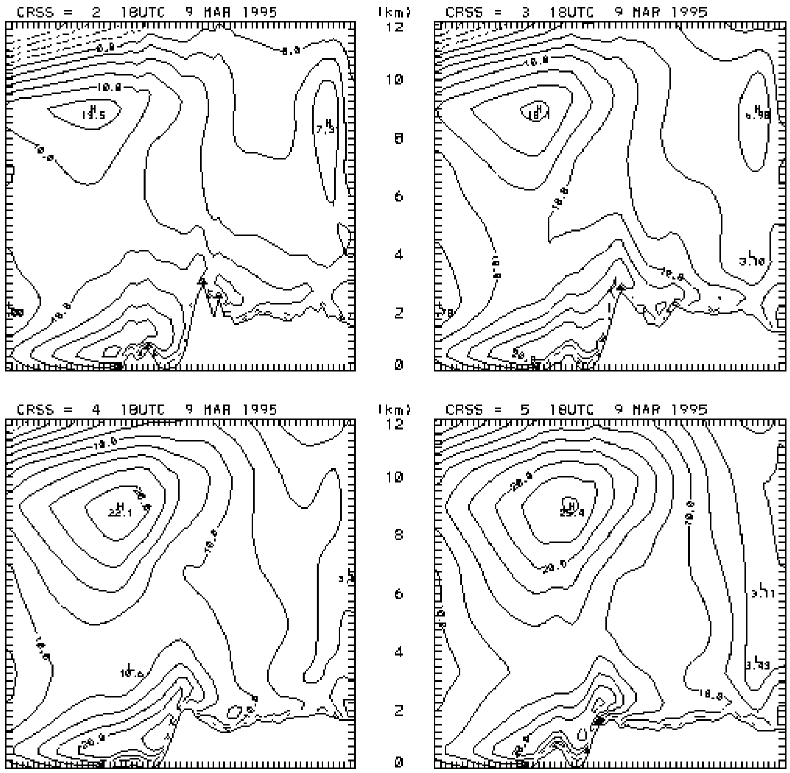


Figure 7. Cross-sections of the along-barrier wind component (the v component in Figure 6) for the mean terrain simulation at 1800UTC on March 9, 1995.

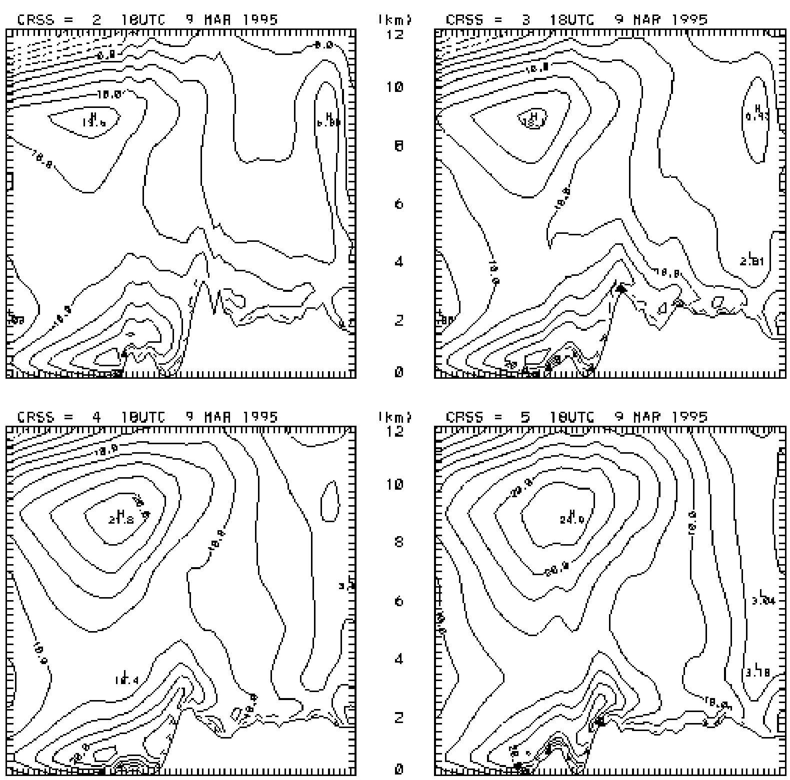


Figure 8. Cross-sections of the along-barrier wind component (the v component in Figure 6) for the envelope terrain simulation at 1800UTC on March 9, 1995.

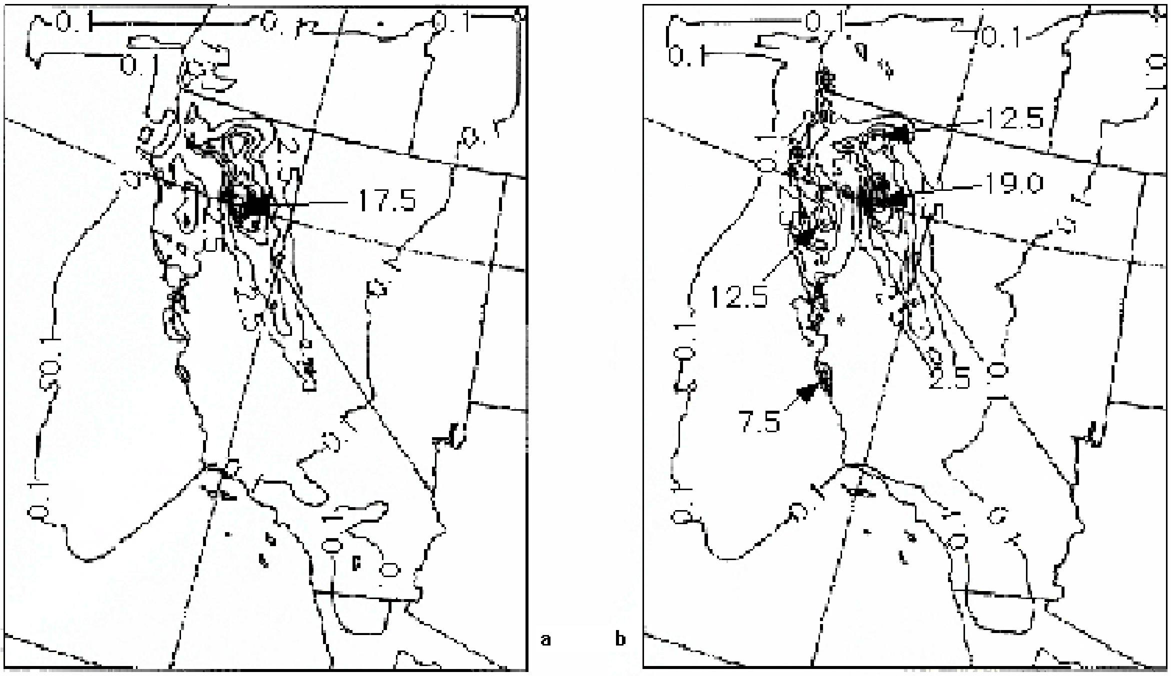


Figure 9. Precipitation isolines at 1800UTC on March 9, 1995 for (a) the mean terrain simulation and (b) the envelope terrain simulation.

CONCLUSIONS

While this work is at a preliminary stage, the results shown here do indicate interesting sensitivities to the details of the representation of the elevation surface. The small-scale simulations presented here suggest that terrain smoothness can have significant effects on the general flow in addition to the direct effect of slowing the winds in direct contact with the terrain. The unsmoothed, rougher terrain provides a resistance to the flow that has a distinct effect that could have important implications for the application of such models to dispersion predictions. That such an effect occurs is reasonable, however, it does raise the question of how determine the degree of smoothness necessary to best match reality and also returns to the question of how to parameterize both subgrid scale terrain variation as well as variation that must be omitted to maintain model stability. In future work, we will attempt to confirm the pattern indicated here by developing more comparison runs that reflect different degrees of smoothness. Also, we will attempt to match the results with field experiment data to determine the most appropriate choice of terrain representation for this type of problem.

The simulated low-level wind fields along with the distribution of precipitation show significant dependence on the terrain representation used in the simulations. When the envelope terrain was used for California, the simulated barrier jet was intensified by over 2.5 ms-1. The envelope terrain also enhanced the low-level vertical motion by 10-30%. This intensified barrier jet and low-level vertical motion enhanced precipitation in northern Coastal Range and Sierra Nevadas and may be very important in achieving accurate precipitation forecasts that can be used to drive hydrologic models of crucial watersheds.

Understanding how the characteristics of geographic data affect their use in complex applications, such as atmospheric modeling, falls within the realm of geographic information science. As discussed by Goodchild (1992), geographic information science examines the unique features of spatial data and the most effective ways to analyze and utilize such data. This work constitutes a step in the direction of understanding the use of elevation data in atmospheric models and future work will be needed to deepen this understanding of elevation data and to broaden into other important types of geographic data that are necessary for mesoscale atmospheric models.

REFERENCES

1. Arakawa, A. and M. Suarez, 1983: Vertical differencing of the primitive equation in sigma coordinates. *Monthly Weather Review*, 111, pp. 34-45.
2. Boybeyi, A., Bacon, D.P., Dunn, T.J., Ho, Y-L., McCorcle, M.D., Peckham, S.E., Sarma, R.A., Young S., and Zack. J. (1994) The Operational Multi-scale Environment model with Grid Adaptivity (OMEGA). Part II.

Simulations of local circulations, *Proceedings of the Tenth Conference on Numerical Weather Prediction, American Meteorological Society*, pp. 369-371.

3. Cho, H.-R., Niewiadomski, J. Iribarne and O. Melo (1989) A model of the effect of cumulus clouds on the redistribution and transport of pollutants. *Journal of Geophysical Research*, 94, D10, pp. 12895-12910.
4. Clements, W.E., Archuleta, J.A., and Gudiksen, P.H. (1989), Experimental design of the 1984 ASCOT field study, *Journal of Applied Meteorology*, 28, pp. 405-413.
5. Deardorff, J. W., (1978) Efficient prediction of ground surface temperature and moisture with inclusion of a layer of vegetation, *Journal of Geophysical Research*, 83, C4, pp. 1889-2330.
6. Georgelin, M., Richard, E., Pettididdier, M., and Druilet, A. (1994) Impact of subgrid-scale orography parameterization on the simulation of orographic flows, *Monthly Weather Review*, 122, pp. 1509-1522.
7. Goodchild, M.F. (1992), Geographical information science, *International Journal of Geographical Information Systems*, 1, pp. 327-334.
8. Hashvardhan, R. Davis, D. A. Randall, and T. Corsetti (1987) A fast radiation parameterization for atmospheric general circulation models. *Journal of Geophysical Research*, D1, 1009-1016.
9. Hsu, Y. and A. Arakawa (1990) Numerical modeling of the atmosphere with an isentropicvertical coordinate. *Monthly Weather Review*, 118, pp. 1933-1959.
10. Kim, J. and S.-T. Soong (1994) Simulation of a precipitation event in the western United States. *Proceedings of the Sixth Conference on Climate Variations*, American Meteorological Society, pp. 403-406.
11. Kim, J. and M. Ek (1995) A simulation of the surface energy budget and soil water content over the Hydrologic Atmospheric Pilot Experiments-Modelisation du Bilan Hydrique forest site. *Journal of Geophysical Research*, 100, D10, pp. 20845-20854.
12. Lee, T.J., Pielke, R.A., Kittel, T.G.F., and Weaver, J.F. (1993) Atmospheric modeling and its spatial representation of land surface characteristics, *Environmental Modeling with GIS*, Oxford, pp. 108-122.
13. Leone, J.M., and Lee, R. (1989) Numerical simulation of drainage flow in Brush Creek, *Journal of Applied Meteorology*, 28, pp. 530-542.
14. Louis, J. F., M. Tiedke, and J. Geleyn (1981) A short history of the operational PBL-parameterization at ECMWF , *Proceedings of the Workshop on planetary boundary parameterization*, European Center for Medium-Range Weather Forecasts, pp. 59-79.
15. Mahrt, L. and H.-L. Pan (1984) A two-layer model of soil hydrology., *Boundary Layer Meteorology*, 29, pp. 1-20.
16. Mason, P.J., (1991) Boundary-layer parameterization in heterogeneous terrain, *Proceedings of the workshop on Fine-Scale Modeling and the Development on Parameterization Schemes*, European Center for Medium-Range Weather Forecasts, pp. 275-288.
17. McNider, R.L., and Pielke, R.A. (1981) Diurnal Boundary Layer Development over Sloping Terrain, *Journal of the Atmospheric Sciences*, 38, pp. 2198-2212.
18. Mesinger, F., Janjic, Z.I., Nickovic, S., Gavrilov, D. and Deaven, D.G. (1988) The Step-Mountain Coordinate: Model Description and Performance for Cases of Alpine Lee Cyclogenesis and for a Case of an Appalachian Redevelopment, *Monthly Weather Review*, 116, pp. 1493-1518.
19. Tibaldi, S. (1986) Envelope orography and maintenance of the quasi-stationary circulation in the ECMWF global models, *Advances in Geophysics*, 29, pp. 339-373.
20. Pielke, R.A. (1984) *Mesoscale Meteorological Modeling*, Academic Press.
21. Starr, D. and S. Cox (1985) Cirrus clouds. Part I: A cirrus cloud model., *Journal of the Atmospheric Sciences*, 42, pp. 2663-2681.
22. Stephens, G., 1978: Radiation profiles in extended water clouds. II: Parameterization schemes., *Journal of the Atmospheric Sciences*., 35, pp. 2123-2132.
23. Soong, S.-T. and J. Kim (1995) Simulation of a heavy wintertime precipitation event in California. *Climatic Change*, in press.
24. Takacs, L. L. (1985) A two-step scheme for the advection equation with minimized dissipation and dispersion errors. *Monthly Weather Review*, 113, pp. 1050-1065.
25. Walker, H., and Leone, J.M. (1994) The impact of elevation data representation on nocturnal drainage wind simulations, *Proceedings of the Sixth Conference on Mesoscale Processes*, American Meteorological Society, pp. 544-547.
26. Wallace, J.M., Tibaldi, S. and Simmons, A.J. (1983) Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography, *Quarterly Journal of the Royal Meteorological Society*, 109, pp. 683-717.
27. Zhong, S., Leone, J.M., and Takle, E.S. (1991) Interaction of the sea breeze with a river breeze in an area of complex coastal heating, *Boundary-Layer Meteorology*, 56, pp. 101-139.

Work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W-7405-Eng-48

Kremer, Robert G., R. B. Lammers, M. Hartman, and J.S. Baron, Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523-1499, U.S.A., and Department of Geography, University of Toronto, Toronto, Ontario, M5S 1A1, Canada.

Biophysical data integration for terrestrial ecosystem simulation from watershed to basin scales.

Abstract

Integration of spatial biophysical data to drive terrestrial ecosystem models requires not only that the data layers themselves be spatially compatible, but also that they adequately resolve landscape attributes that the models require. We have integrated soils, topographic, vegetation, and meteorological data to simulate carbon, nitrogen, and water dynamics at watershed (7 km) to basin (63,000 km) scales within the South Platte River Basin in Colorado, Wyoming, and Nebraska, using varying data resolutions and levels of topographic aggregation. Simulations indicate that spatial correlation of input variables with synoptic weather patterns is critical in determining dynamic processes such as hydrologic outflow and evapotranspiration, yet is not always possible across large regions or heterogeneous terrain. We examine the flow of spatial biophysical data as input to multiple-scale ecosystem process simulations and discuss how spatial data requirements, compatibility, and availability are affected across scales.

Simulation of ecological processes at regional to global scales typically requires both intensive and extensive biophysical parameterization of the earth's surface and atmosphere. Unfortunately for ecological modelers, critical processes are continuously interacting at spatial scales from local (e.g. watershed hydrologic partitioning) to regional (e.g. synoptic meteorological events), and at temporal scales that range from instantaneous (e.g. vapor flux densities) to seasonal (e.g. vegetation phenology), and are thus not easily incorporated into ecosystem models that require multiple data of fixed spatial or temporal resolution as input. In order for ecological models to adequately describe or predict physical processes such as net primary productivity or streamflow dynamics, input data resolution must be compatible not only with the scale of actual ecological functioning, but also with the resolution at which interpretation of results is critical.

The compilation and implementation of spatial biophysical data into ecosystem process simulation has at least four basic considerations. 1) What are the minimal biophysical data required to parameterize all relevant processes in the model? 2) What are the maximum spatial and temporal resolutions of these data that will render meaningful simulations? That is, what are the spatial and temporal levels at which the processes are critical in the real, not simulated, world? 3) Do these data exist, or can they be formulated from existing data? 4) Once required or optimal data resolutions are determined or hypothesized, are multiple data "layers" compatible in their resolutions to be combined in a geographic information system?

As we move toward the application of ecosystem models to continental and global scales, it is essential to aggregate small-scale ecologic functioning to scales that allow current computational hardware to store and process the data without insurmountable time and costs involved. In aggregating spatial data to these scales, however, critical ecological, climatic, or topographic detail may be lost, and large scale simulation results may become an aggregation of errors rather than broad representation of earth processes.

For example, through modeling studies that utilize suites of spatial data to simulate landscape ecological processes, we have tested the effects of partitioning landscapes from sub-watershed to regional scales on hydrologic and ecosystem process model simulations . Results showed that as we decrease the scale of topographic resolution to match coarse resolution biophysical data input to the models (e.g. 1 km AVHRR data for vegetation characteristics), key topographic relationships with primary production, meteorology, and hydrologic partitioning are lost, and simulation errors dramatically increase across regional and larger scales. Lammers, et al. (1995) state, "Initial attempts to extend these models over larger land areas implicitly assumed that the surface could be treated as a spatially exhaustive set of homogeneous areas, acting independently and in parallel. However, recent work has demonstrated that surface heterogeneity is both strongly expressed at all scales and is not simply averaged in a functional sense (Avissar, 1992; Band, 1993). The nonlinear response of water and carbon flux processes to available soil water, meteorological variables and certain vegetation canopy attributes commonly results in significant bias when computing areal averaged flux using mean or average surface conditions... In scaling or aggregating a biophysical model over progressively larger areas, a key problem is estimation of this distributional information as direct sampling becomes infeasible."

In one study (Lammers et al., 1995) simulations of forest ecosystem processes over a 3000 km*2* watershed were used to investigate scale effects on sampling and representing land surface attributes. Specifically, the development and control of bias in simulated carbon and water exchange processes as both scale and resolution of the landscape changed was investigated. This study showed that an order of magnitude resolution change of the original land data sets for topography and vegetation cover can produce similar results in the carbon and water flux processes as long as a joint distribution function describing significant surface attributes is preserved. It may be possible to define this distribution function by partitioning surface heterogeneity as variance between the spatial units used for simulation and variance within units. This scheme implies, however, that biophysical attribute data are available at a common, minimum resolution, and that within-unit variance may be tracked across aggregate scales. When coupling multiple "layers" in a GIS, data are typically gathered from disparate sources, and common scale must be initially reconciled, which in itself, ideally, requires tracking of variance-to-scale relationships. In another study (Band et al., 1995), the effects of digital elevation model (DEM) resolution were investigated for the computation of hydrological terrain- soils indices for input to hydrological models. This study found that DEM resolution appears to have regular impacts on simulated hydrographs, with greatest sensitivity in grid based (raster) methods and the least sensitivity in contour and slope line based (vector) methods. Again, this problem may be confounded when varying degrees of topographic representation must be coupled with other biophysical data such as vegetation quantification (leaf area index, specific leaf area, etc.). Finding mutual distribution functions among various layers in a GIS may be a formidable task, but one that would ultimately define the interaction of topography-vegetation-soils-hydrology.

Whether we are simulating local or global scales, physical processes must be described at the scale at which they are critical to ecological functioning. As a scientific community, however, we are currently severely limited in our ability to adapt input databases to address specific ecological questions because of inherent problems and variability of biophysical information across scales. Rather we must spend our time adapting arbitrarily-scaled data sets to one another, all the while trying to evaluate what we are missing in the first place with coarse resolution data, and what we are losing in the process of scaling fine-resolution data to coarser scales. Quite often databases are created with resolutions borne of hardware limitations and processing times, rather than regard to relevant physical scales. Because it is always better to scale fine resolution data up to larger scale applications than the converse (i.e. loss of feature or process resolution can be monitored and evaluated), databases should be created at the finest resolution possible, from the initial hardware (e.g. satellites) to archiving and distribution, so that end users of the data can objectively evaluate the level of aggregation that is suitable to the specific application.

The objective of this study is to illustrate the need for spatially compatible biophysical data sets by presenting examples in which scale is shown to be a critical parameter in simulation results over a given landscape. By tracing the flow of data through coupled hydrologic-ecosystem process models, we identify points at which data layer resolution incompatibility becomes a problem in addressing the ecological questions the models are designed to answer.

References

Lammers, R.B., L.E. Band and C.L. Tague (1995) Scaling Behavior of Watershed Processes. In P. van Gardingen, G. Foody and P. Curran (eds.) Scaling-up. Final revision submitted, Cambridge University Press.

Avissar, R. (1992) Conceptual aspects of a statistical dynamical approach to represent landscape subgrid-scale heterogeneities in atmospheric models. Journal of Geophysical Research 97, 2729-2742.

Band, L.E. (1993) Effect of land surface representation on forest water and carbon budgets. Journal of Hydrology 150, 749-772.

Band, L.E., R. Vertessy and R.B. Lammers (1995) The Effect of Different Terrain Representations and Resolution on Simulated Watershed Processes, Zeitschrift fur Geomorphologie, vol. 101, November, pp. 187-199.

Robert G. Kremer

Natural Resource Ecology Laboratory

Colorado State University

Fort Collins, CO 80523

970/491-1966

FAX:970/491-1965

rkremer@nrel.colostate.edu

Fill in your name and email address here

Lin Wu

An Integration of a Surface Energy Balance Climate Model with TIN and GRID in GIS

Abstract.

A surface energy balance model simulates surface temperature and sensible, latent, and subsurface heat fluxes of the boundary layer, the atmospheric layer near the ground level. Equations of a surface energy balance model are usually established based on the assumption that the surface in the modeling area is homogeneous in radiative, thermal, and geometrical properties. In case of a nonhomogeneous area, such as a hilly area or an urban area, the surface has to be partitioned into small, homogeneous facets so that the equations can be applied to each uniform facet. Surface partitioning has been a difficult task in surface energy balance model establishment. Since the GRID and TIN structure in a GIS decompose surfaces into relatively homogeneous facets, the integration of a surface energy balance model with GRID or TIN will enhance the model's ability to simulate the spatial variation of surface temperature and heat fluxes in heterogeneous areas. The integration also enhances the model's applicability in applied fields such as agriculture and urban and regional planning. The paper examines the theoretic basis of surface energy balance models, different input data format from GIS, the characteristics of the integrated models, the format of the simulated results, and the advantages and limitations related to the integration.

Introduction

Surface energy balance models simulate microscale energy exchange processes between the ground surface and the atmospheric layer near the ground level (Oke, 1987). These processes include radiative, sensible heat, latent heat, and subsurface heat exchange processes. Results from the model provide climatic information such as surface temperature, radiation and heat fluxes related to particular surfaces. This type of information, often lack in a GIS, is important to studies of vegetation, agriculture, urban design and urban planning (Oke, 1982).

The surface energy balance model used in this study was established based on the principles of energy balance equations (Bornstein, 1986; Oke, 1987; Terjung, 1978) which were developed based on the assumption that the surface in the modeling area is homogeneous in radiative, thermal, and geometrical properties (Carlson et al., 1981; Outcalt, 1971; 1972). In case of a nonhomogeneous area, such as a hilly area, or a partially vegetated area, the surface has to be partitioned into small, homogeneous facets so that the equations can be applied to each uniform facet (Terjung, 1978). Surface partitioning has been a difficult task in surface energy balance model establishment. This difficulty became one of the obstacles for model development and application.

Since GRID and TIN structure in a GIS decompose surfaces into relatively homogeneous facets, the integration of a surface energy balance model with GRID or TIN will enhance the model's ability to simulate the spatial variation of surface temperature and heat fluxes in heterogeneous areas. The results from the model in a GIS can be easily classified or indexed with the spatial analysis functions of a GIS, and displayed spatially, which significantly improve the model's applicability.

This research is an attempt to embed an energy balance climate model into a GIS based on the existing GIS analysis capabilities and macro language functions (Goodchild, 1993).

Principles and Guiding Equations of the Surface Energy Balance Model

The development of the surface energy balance model is based on the assumption that the temperature and energy exchange processes of a surface are governed by the basic energy balance equation:

$$Q^* = K^* + L^* = Q_H + Q_E + Q_G \qquad (Wm^{-2}) \qquad (1)$$

where Q* is the net all-wave radiation which is the balance between the net shortwave radiation (K*) and net longwave radiation (L*). The net radiation input is balanced by three heat fluxes: sensible heat flux (QH), latent heat flux (QE), and subsurface heat flux (QG). The equation is solved by searching a unique equilibrium surface temperature that results in a balance.

Shortwave Radiation

The net shortwave radiation (K*), in Equation (1), is the balance between the incoming shortwave radiation (*Kdown*) and the reflected shortwave radiation (*Kup*):

$$K^* = K\downarrow - K\uparrow \qquad (Wm^{-2}). \tag{2}$$

On an opaque surface, there is no radiative transmission through the surface, therefore, with given albedo (*a*) or absorptivity (*zetashort)*, the reflected shortwave radiation (*Kup*) in Equation (2) can be derived based on incoming shortwave radiation (*Kdown*),

$$K\uparrow = \alpha K\downarrow = (1 - \zeta_{short})K\downarrow \qquad (Wm^{-2}). \tag{3}$$

Kdown involves two major components, direct solar radiation (S) and diffuse solar radiation (D),

$$K\downarrow = S + D \qquad (Wm^{-2}). \tag{4}$$

Direct solar radiation (S) can be represented by,

$$S = S_0 \cos\theta \qquad (Wm^{-2}) \tag{5}$$

where *S0* is direct radiation at normal incidences, and is a parameter determined by slope angle (*phi*), slope azimuth (*OMEGA sl*), zenith angle (Z), and solar azimuth (*OMEGA s*):

$$\cos\theta = \cos\varphi\cos Z + \sin\varphi\sin Z\cos(\Omega_{sl} - \Omega_s) \tag{5a}$$

Between sunrise and sunset, the zenith angle is given by:

$$\cos Z = \sin\phi\sin\delta + \cos\phi\cos\delta\cosh \tag{5b}$$

$$h_{rise} < h < h_{set}$$

where *PHI* , *delta* , and h are latitude, declination angle, and hour angle. Declination angle (*delta*) is a dependent variable of the date of the year and is derived from an array of values. The hour angle h defines the time of a day. By setting the zenith angle to 90, the sunrise and sunset hour angles (*hrise*, *hset*) can be derived with Equation (5b).

The solar azimuth angle *OMEGA s* is a function of latitude, declination angle, and zenith angle at a given time:

$$\cos\Omega = \frac{\sin\phi\, cosZ - \sin\delta}{\cos\phi\, sinZ} \tag{5c}$$

$$\Omega_s = 0, \quad if \quad |\cos\Omega|>0; \quad \Omega_s = 2\pi - \Omega, \; before 1200; \quad \Omega_s = \Omega, \; after 1200.$$

S0 in Equation (5) is determined by,

$$S_0 = I_0 V^2 \Psi \qquad (Wm^{-2}) \tag{6}$$

where *I0* is the solar irradiance (*I0 = 1370 Wm-2*) (Kale, 1985), V represents radius vector which adjusts the distance between the earth and the sun, and is the transmission coefficient which is a function of atmospheric pressure (p), precipitable water vapor (w), optical air mass (m), and dust-haze factor (d) (Brooks 1959),

$$\Psi = \exp[-0.089(\frac{pm}{p_0})^{0.75} - 0.174(\frac{wm}{20})^{0.60} - 0.083(dm)^{0.90}] \qquad (6a)$$

$$m = \sec(z)\frac{p}{p_0}; \qquad w = 10^{(0.247\sqrt{e_a} + 0.421)},$$

where p0 is sea level pressure (p0 = 101.3 kPa). The dust-haze factor d has a value ranging from 0.6 to 1.0 for "clear" conditions, and from 1.4 to 2.0 for polluted conditions (Terjung, 1978). The optical air mass m is adjusted by the elevation z, p is the pressure at elevation z. The precipitable water vapor w varies with vapor pressure, ea.

Diffuse solar radiation (D) comes from two sources: the diffuse radiation from the sky, and the reflected radiation from the environment. Because obtaining reflected radiation from the environment is a very complicated process, a separate study of the integration between this part of the model and a GIS is required. In this study, the model is simplified to assume that the diffuse radiation comes from the sky only. Since the equations in this study are applied to relatively small facets (cells in GRID and triangles or small irregular facets in TIN), when there is no drastic change of the terrain, this assumption could be true (the facets don't "see" their neighboring facets, therefore they don't receive the reflection from the neighbors). Under such an assumption, the diffuse radiation can be calculated by,

$$D = 57.24 + 0.1253S_s - 0.1421S_h - 17.16m + 231.7\log(\frac{w}{10}) \quad (Wm^{-2}) \qquad (7)$$

where Ss and Sh are direct radiation on the top of the atmosphere and on a horizontal surface at the bottom of the atmosphere (both can be derived from equation (5) and (6)), m is the optical air mass, and w is the precipitable water vapor. Equation (7) is based on an empirical approach developed by Terjung and Louie (1973) from observations in the Los Angeles area. Because this equation was developed for Los Angeles, adjustments may be necessary for other locations. An isotropic condition is assumed for the diffuse radiation from the sky.

All the variables in Equation (2) are either given or be solved yielding the shortwave radiation budget components for Equation (1).

Longwave Radiation

The net longwave radiation is determined by the absorbed incoming longwave radiation, L and outgoing longwave radiation L :

$$L^* = \zeta_{long}L\downarrow - L\uparrow \qquad (Wm^{-2}) \qquad (8)$$

where long is the longwave absorptivity of the surface.

The outgoing longwave radiation from a surface (*Lup*) can be determined by the Stefan-Boltzmann law:

$$L\uparrow = \sigma\epsilon_s T_s^4 \qquad (Wm^{-2}) \qquad (9)$$

where is the Stefan-Boltzmann constant (*5.67x10-8 Wm-2K-4*), s is the emissivity of the surface given as a surface property, and Ts is the surface temperature in Kelvin.

The incoming longwave radiation comes from two sources: the sky and the environment. Similar to the diffuse radiation, it is assumed in this study that the incoming longwave radiation comes from the sky only:

$$L\downarrow_{sky} = \sigma\epsilon_a T_a^4 \qquad (Wm^{-2}) \qquad (10)$$

$$\epsilon_a = (0.66 + 0.039\sqrt{e_a}),$$

where Ta is the shelter height air temperature and ea is the vapor pressure. This formula was developed by Brunt (1932), which, compared to other equations, was found to be more stable with temperature change than other methods (Jiménez et al, 1987).

All the variables required to solve longwave radiation are known except surface temperature (Ts) in Equation (9) which will be discussed later.

Heat Fluxes

Heat fluxes are the responding processes to the changing energy input from net radiation. Q* is partitioned into sensible heat flux (QH), latent heat flux (QE), and subsurface heat flux (QG) (Equation (1)).

Sensible heat flux QH is expressed as a function of the temperature difference between the surface and the adjacent air and a heat transfer coefficient (hc)(Terjung and O'Rourke, 1980):

$$Q_H = h_c\ (T_s - T_a) \qquad (Wm^{-2}), \qquad (11)$$

$$h_c = 2.84(0.00033u)^{0.8} \qquad (Wm^{-2}\ K^{-1})$$

where u is the wind speed measured at shelter height (*m s-1*). Since Ta in Equation (11) is given by input data, the only unknown variable in Equation (11) is the surface temperature (Ts).

To derive latent heat flux (QE), a Bowen Ratio approach (Oke, 1987) is used. Bowen Ratio (*beta*) is the ratio between sensible heat flux and latent heat flux. Typical values of *beta* range from less than 1 for very humid vegetated environments to greater than 10 for very dry desert environments (Oke, 1987). Based on Equations (11), and with given *beta* , QE can be presented as:

$$Q_E = \frac{Q_H}{\beta} = \frac{1}{\beta} h_c (T_s - T_a) \qquad (Wm^{-2}). \qquad (12)$$

All the variables in Equation (12) are known except Ts .

The last term in Equation (1) is the subsurface heat flux (QG), the energy exchange between the surface and the subsurface. QG is determined by the subsurface temperature gradient and thermal conductivity of the substrate (k):

$$Q_G = k \frac{T_s - T_g}{d} \qquad (Wm^{-2}). \qquad (13)$$

Tg is the subsurface temperature at a distance (d) below the surface and is given as input data. k is given based on surface material and moisture availability (Oke, 1987). The only unknown variable in Equation (13) is, again, the surface temperature Ts.

Surface Temperature

If we examine each term in Equation (1) again, we can see that the first term, net shortwave radiation K*, is solved by Equation (2) through (7) and can be viewed as a constant. The second term, net longwave radiation L*, can be represented by Equations (8) through (10), in which the only unknown variable is the surface temperature Ts. Therefore, L* can be viewed as a function of Ts, L* (Ts). The sensible heat flux (QH), latent heat flux (QE), and subsurface heat flux (QG), all include one common unknown variable: surface temperature Ts and each can be represented as functions of Ts: QH(Ts), QE(Ts), and QG(Ts) respectively. Substituting each term in Equation (1) we have:

$$K^* + L^*(T_s) + Q_H(T_s) + Q_E(T_s) + Q_G(T_s) = 0 \qquad (Wm^{-2}) \qquad (14)$$

Surface temperature (Ts) is obtained by an interactive numerical approach (Gerald and Wheatley, 1994). The processes begin with an initial Ts value. Each term is solved based on the initial value. If the result exceeds zero, Ts is adjusted by a given interval and each function of Ts is recalculated. After every calculation, the interval is reduced by half. The process is repeated until the equation balances when the difference between both sides of the equation falls within a given tolerance level (for example, less than 1.0 *Wm-2*). At this point, the surface temperature is solved and so are L*, QH, QE, and QG.

Integration of the model with GRID and TIN

The model described in the previous section was integrated with GRID and TIN in an ARC/INFO workstation. A GRID is a raster-based system which divides a continuous surface into square cells based on a given cell size (resolution). Each cell possesses a value which indicates one attribute of the cell. Analysis functions based on individual cells (local operation), neighboring cells (focal operation) and all the cells in the database (global operation) are built into the system. Many of the cartographic modeling capabilities (Berry, 1993) and surface analysis capabilities are available in GRID system. A TIN (Triangulated Irregular Network) divides a continuous surface into irregular triangles based on given control points. The principles of Delauney Triangulation (Burrough, 1986) are used to construct the triangles in a TIN. Although conceptually, a TIN can be viewed as a raster system (Burrough, 1986), the model integration with TIN in this research is mainly based on the vector function of the system.

Although the specific procedures in integrating the model with GRID and TIN are different (the former is based on raster procedures and the latter vector procedures), the principles of the integration are identical in both cases. Both GRID and TIN divide a continuous surface into uniform units, square cells in GRID and irregular facets in TIN (Burrough, 1986; ESRI, 1991a, 1991b). It is assumed that no variation occurs across a single unit surface, whether it is a square cell or an irregular triangle or polygon. This provides a foundation for the surface energy balance model because the energy balance can be sought for each individual unit which meets the assumption of homogeneity for the energy balance equations. When simulations are carried out for each and all the cells or facets that compose a nonhomogeneous surface, the spatial variation over the surface is derived.

The integrated model with GRID is illustrated by Figure 1.

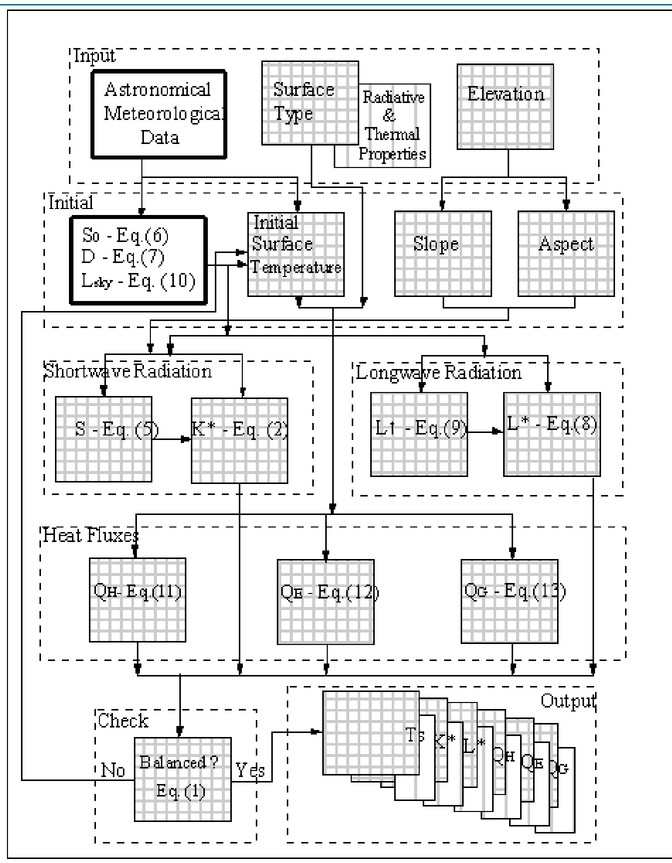


Figure 1 Illustrative flow chart of the integrated model with GRID. Boxes with grids on the background indicate GRID based format or operation, boxes with vertical lines indicate attribute data tables, and boxes with thick boarders indicate macro language blocks.

The input data of the model are given by three different formats. The astronomical data (time, date, and latitude, etc.) and the non-surface dependent meteorological data (input air temperature, pressure, etc.) are given by assigning respective values to system wide variables in macro language blocks. The spatially varied input data (elevation and surface type) are given by GRID format. The gird for the surface type can be created based on classified surface types (eg. vegetated surface, bare soil surface, paved surface, etc.) or based on classified aerial photographs. The radiative and thermal properties associated with each surface type are given by attribute tables related to the surface grid. The initialization stage of the simulation involves further processing of the input data. The grid layers representing slope and aspect are created from the elevation layer with existing GRID functions. Incoming longwave radiation, diffuse radiation and the direct radiation at normal incidences are calculated based on input data by macro language blocks. An initial surface temperature grid is created based on either the subsurface temperature or the air temperature (the initial value will be replaced by simulated Ts values). After the initialization, simulations of the radiation terms and the heat fluxes terms are carried out based on the initial values with existing grid functions. The energy balance equations are solved for every grid in the study area. The energy balance for each of the cells is checked. The cells that reached the balance are recorded to an output layer and a mask is created to block out these cells for further processing. For the unbalanced cells, the initial surface temperatures are changed based on a given interval and the simulations are repeated until all the cells reach their balance. The results are then written to several output grid layers. From there, further classification and conversion to vector format are possible.

The integration of the model with TIN is illustrated by Figure 2.

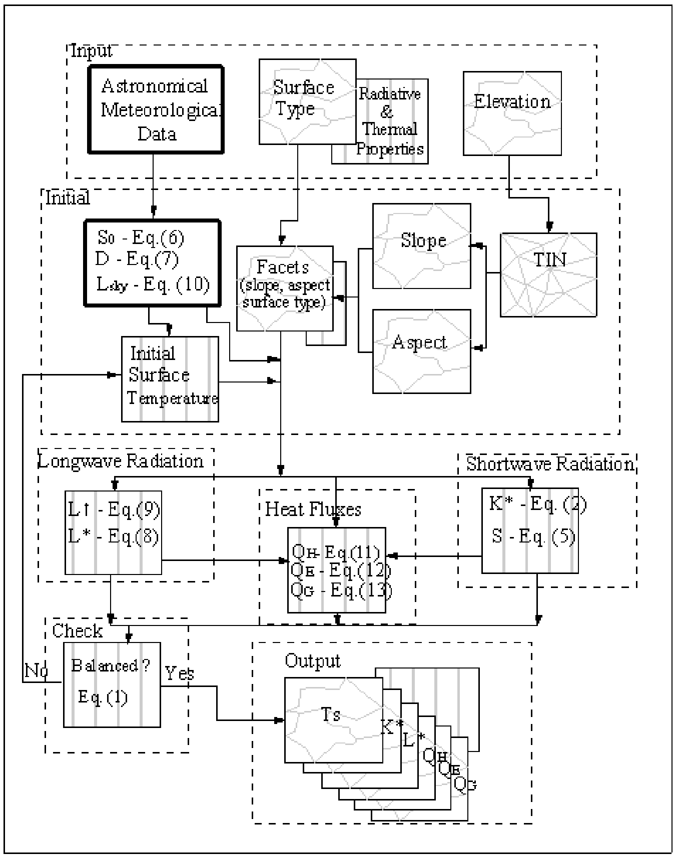


Figure 2 Illustrative flow chart of the integrated model with GRID. Boxes with polygons on the background indicate vector based format or operation, boxes with vertical lines indicate attribute data tables, boxes with triangles on the background indicate TIN structure, and boxes with thick boarders indicate macro language

The input component of the integrated model with TIN is identical to the GRID except that the spatially varied input data now are entered with polygon layers. At the initial stage, a TIN is generated from the elevation layer. Because TIN is not directly associated with any attribute tables, it has to be converted back to polygon layers (all the polygons are triangles) so that the slope and aspect information can be retained in attribute tables associated with the polygon layers. The slope and the aspect layers are then combined with the surface type layer by the existing vector analysis function of the GIS to create a polygon layer composed of uniform facets (no longer triangles because the combination with the surface type) in terms of its slope, aspect, and surface types. From this point on, almost all the simulations are carried out in attribute format using the facets as the basic units.

Conclusions

The study suggested that the surface partitioning capabilities and the spatial operation functions of a GIS provide a surface energy balance climate model with a framework to derive the spatial variation patterns of various climate parameters. The integration is mainly based on existing functions of the GIS and it macro language capabilities. By integrating with a GIS, the model is improved in the following aspect:

a. The model is supplied with spatially varied input data which enhance the model's capability of providing spatial patterns;

b. The cell or small polygon-based simulation unit ensures that the assumption of homogeneity for the energy balance equations is met more precisely; and

c. The spatial analysis function of the GIS enables the model's output to be displayed spatially and to be classified in various ways which greatly enhances the results analyses process and application.

The integrated model at this stage still has limitations. As discussed earlier, the model is simplified in some aspect to make the integration possible. One of the simplifications is the elimination of the radiative input from the environment - the neighboring cells or facets. Such a simplification will certainly affect the model's accuracy under certain conditions when longwave radiation and diffuse radiation from the environment are important factors. Since simulation of the environmental effects is a complicated process (Johnson and Watson, 1984; Oke, 1987; Steyn, 1980) existing GIS functions at current level cannot provide easy solutions. Further research is needed to study the possibility of this part of the integration.

Acknowledgments

This research is supported by the Research, Scholarship and Creative Activity Program at California State Polytechnic University, Pomona.

References

Berry, J. K., (1993) Cartographic Modeling: The analytical capabilities of GIS. Environmental Modeling with GIS. Oxford Univ. Press. pp. 58-74.

Bornstein, R. D., (1986) Urban climate models: nature, limitations and applications. WMO No.652, pp 237-276.

Brooks, F. A., (1959) *An Introduction to Physical Microclimatology* Davis: Univ. of Calif.

Brunt, D., (1932) Notes on radiation in the atmosphere. *Quart. J. Roy. Meteor. Soc*., 58: 389-418.

Burrough, P.A., (1986) *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford Univ. Press.

Carlson, T. N., Dodd, J. K., Benjamin, S. G. and Copper, J. N., (1981) Satellite estimation of the surface energy balance, moisture availability and thermal inertia. *Jour. of Climate and Appl. Meteor*., 20: 67-87.

Gerald, C. F. and Wheatley, P. O. (1994) *Applied numerical analysis*., 5th ed. Addison-Wesley Pub.

Goodchild, M. F., (1993) The state of GIS for environmental problem-solving. *Environmental Modeling with GIS*. Oxford Univ. Press. Pp 8-15.

ESRI, (1991a) *Cell based modeling with GRID*. ESRI

ESRI, (1991b) *Surface modeling with TIN*. ESRI

Jiménez, J. I. , Alados-Arboledas, L., Castro-Diez, Y. and Ballester, G., (1987) On the estimation of long-wave radiation flux from clear skies. *Theor. Appl. Climatol*., 38: 37-42.

Johnson G. T. and Watson I. D., (1984) The determination of view-factor in an urban canyon. *Jour. of Climate and Appl. Meteor*., 23: 329-335.

Kyle, H. L., Ardanuy, P. E. and Hurley, E. J., (1985) The status of the Nimbus-7 earth-radiation budget data set. *Bull. Am. Meteor. Soc.*, 66: 1378-1387.

Oke, T. R., (1986) Urban climatology and the tropical city: an introduction. in *Proc. of the technical conference of urban climatology and its applications with special regard to tropical areas*, WMO No.652, pp1-25.

Oke, T. R., (1987) *Boundary layer climates*. 2nd ed., Methuen, N.Y.

Outcalt, S. I., (1971) A numerical surface climate simulator. *Geog. Analysis*, 3: 380-303.

Outcalt, S. I., (1972) The development and application of a simple digital surface-climate simulator. *Jour. Appl. Meteor*., 11: 629-636.

Steyn, D. G. (1980) The calculation of view factors from fish-eye lens photographs. *Atmosphere-Ocean*, 18: 254-258.

Terjung, W. H. and Louie, S. S., (1974) A climatic model of urban energy budgets. *Geog. Analysis*, 6: 341-367.

Terjung, W. H. and O'Rourke, P. A., (1978) *An outline of boundary-layer climatology methods and analysis.* Academic Publishing Service ASUCLA, Los Angeles.

Terjung, W. H. and O'Rourke, P. A., (1980) *Energy exchanges in urban landscapes: selected climatic models.* Publ. in Clim., Vol. XXXIII, C. W. Thornthwaite associates..

Lin Wu

Assistant Professor

Department of Geography and Anthropology

California State Polytechnic University, Pomona

3801 W Temple Ave

Pomona, CA91768

Telephone: (909)869-3578

Fax (909)869-3586

E-mail: LWU@CSUPomona.edu

THE POTENTIAL OF GIS AND COUPLED GIS/CONVENTIONAL SYSTEMS TO MODEL ACID DEPOSITION OF SULPHUR DIOXIDE

U. Dragosits, C.J. Place and R.I. Smith

Abstract - Research into atmospheric pollution is of vital importance for man's future welfare. In particular, there is a need to explain the causal connections and model the processes involved. Towards this end, GIS have been explored as an alternative to more traditional methodologies using 3rd generation programming languages. However, publications in this field have largely been confined to qualitative statements. This research aims to add to the existing knowledge base by implementing a sulphur dioxide deposition model within both a "pure" GIS system and an integrated GIS/conventional environment. In contrast to previous work, both quantitative and qualitative results are presented. Experiments were carried out for the area of Halladale in northern Scotland.

The results indicate that the implementation of the model within the GIS (ARC/INFO) is approximately two orders of magnitude slower than the coupled GIS/conventional implementation, with equivalent numerical accuracies. The integrated environment was found to incorporate the benefits of both the fast iterative capabilities of a 3rd generation language and the advantages of the GIS such as ready coded spatial representation, flexibility, integration of other data sets and post-iterative analysis. The lessons learned from this study can be applied to other environmental models with a highly iterative element.

1. Introduction

Air pollution is a major concern today. It is therefore vital to investigate the processes and mechanisms contributing to this global environmental problem. In order to further this investigation, models which attempt to represent the processes of emission, atmospheric transport, chemical transformation and/or deposition through mathematical and statistical approximations have been developed. Since the processes involved in air pollution have a spatial component, it would be expected that GIS, which inherently use space as their governing concept (ARONOFF, 1991), should prove well suited. GIS and environmental modelling is presently one of the most topical issues in current GIS research (see FEDRA 1993a, b; KEMP 1992; NYERGES 1992). There is a multitude of air pollution models using GIS (COLLINS et al., 1994; TROZZI and VACCARO, 1993; RINALDI et al., 1993; LEE and PIELKE, 1993; REYES et al., 1993; STEINNOCHER and KNOFLACHER, 1994). However, this discussion does not seem to have an impact on the mainstream air pollution literature, e.g. the journal "Atmospheric Environment".

It is therefore important to integrate the knowledge of GIS specialists and air pollution specialists to a greater degree. As a means to this end, this study uses a model of the acid deposition of sulphur dioxide (SO2) to investigate to what extent geographical information systems (GIS) can contribute to the ongoing development of such models. The emphasis was on comparing and benchmarking two different approaches - a "GIS only" and a coupled GIS/conventional systems implementation using the same data and modelling equations. Experiments were carried out in order to quantify and qualify the advantages and disadvantages of both approaches.

In the subsequent paragraphs, the following aspects will be covered:

- a brief introduction into spatial environmental modelling
- a short description of the applied model for the deposition of SO2 to the surface
- the description of the study area and the data available for input into the model
- a brief specification of the hardware and software configuration
- a description of the implementation of the model, comparing experimentally several versions, implemented inside a GIS and in coupled systems environments
- the discussion and evaluation of the results, weighing the advantages and disadvantages of the GIS- and the coupled systems implementations against each other.

2. Environmental modelling and air pollution modelling using GIS

Environmental systems are among the most complex in science. In order to understand such systems, assess the human impact on them, predict future development and design management strategies, they are often abstracted into (simplified) mathematical or statistical models.

In atmospheric modelling, the relationships between complex processes and space - topography, land cover, vegetation parameters and meteorological conditions, which are all spatially distributed - are evident. Within the domain of environmental modelling, these relationships have more recently been addressed by the use of spatially distributed models that describe environmental processes and phenomena in (continuous) space instead of simple spatially aggregated or lumped parameter models (MAIDMENT, 1993, p.4). This trend is partially driven by the more powerful and affordable technology available today, which also eases the implementation of linked or integrated system environments (FEDRA, 1993a).

The purposes and - at the same time - benefits of combining environmental modelling and GIS are many:

- to describe, understand and explain the system behaviour in a spatial/landscape context;
- as an aid to thought, communication and experimentation (SHANNON, 1975);
- once the governing principles of a system are established, the model can be used
  - to predict results where measurement is impossible and/or too expensive,
  - to test "what-if" scenarios (such as the impacts of an increase in atmospheric SO2 concentrations or changing land cover on acid deposition), or
  - to calculate best/worst case scenarios.
- The results of spatial environmental modelling can be a very valuable contribution to decision support in planning, e.g. the development of efficient and cost effective air quality control strategies (CARMICHAEL and PETERS, 1984b).

The integration of both methods, GIS and environmental modelling, is apparently of advantage for progress in environmental research, since both methods are synergistic. More generally, no computer system is perfect for all purposes of modelling. Since there is no such thing as a "general purpose" spatial environmental modelling system and all the existing systems have their particular strengths and weaknesses, it might often be required to use more than one system to cover all aspects of an application area. By coupling complementary systems together, the resulting combination should be more efficient (NYERGES, 1992). Although, the more integrated the solution is, the more expensive is its development. "With interest in coupling models to GIS on the rise, due to the maturity of both processing environments and recognising that no software can do it all, we can expect to see more developments in standard interface services in the foreseeable future" (NYERGES, 1992, pp. 542).

Over the last few years, GIS have been used in many atmospheric models, mostly coupled with other systems (TROZZI and VACCARO, 1993; RINALDI et al., 1993; LEE and PIELKE, 1993; REYES et al., 1993; STEINNOCHER and KNOFLACHER, 1994). GIS were used for data input and management, preparation, display and spatial analysis of the results, high quality mapping and user interfaces. All these approaches integrate GIS and other systems to different degrees. FEDRA (1993a, b) collates the different degrees of coupling GIS to other systems for environmental modelling as follows:

- "Loose coupling" is described as sharing files between systems with separate interfaces (relying on manual resolution of differences in the data representation);
- "Tight coupling" as enabling the shared files to be used in either interface. The user interface(s) take care of the data management, but they are still independent interfaces.
- An "integrated approach" consists of shared files and memory in the computer and only one interface.

REYES et al. (1993) argue that, since "commercial technology is at a stage where the use of sophisticated mathematical models has not been fully incorporated" (REYES et al., 1993, pp. 4), the most appropriate solution for their application was to use the GIS as the front end for the interaction between the user with both the GIS and mathematical models. MAIDMENT (1993) takes a different approach, believing that the best solution to the existing problems in spatial environmental modelling is not to link the different existing systems, but rather to find a way to change GIS by gradually developing GIS functions for analytical solutions of particular environmental processes. He suggests that this should be approached through a synthesis of knowledge of specialists in the relevant fields by trying to isolate the fundamental processes and building GIS tools to represent them. Which of the two ways to approach the present problems of spatial environmental modelling is best, cannot be answered easily. Therefore it is important to try to quantify the advantages and disadvantages of a "GIS-only" model in an application area such as dry deposition modelling versus a coupled systems approach.. This was one of the aims of this study.

---

3. Acid deposition - governing processes

There are numerous environmental problems resulting from air pollution. One of them is acid deposition, mainly caused by sulphur compounds, oxides of nitrogen (NOx) and ammonia (NH3). The processes taking place in the atmospheric boundary layer from the release of pollutants to their deposition are very complicated and vary continuously according to atmospheric conditions (such as wind speed, temperature, solar radiation etc.) and the state of the earth's surface (e.g. vegetation cover, terrain characteristics) (SMALL and SAMSON, 1989). In the following paragraphs, the current understanding of acid deposition and its approximation through a fairly established deterministic mathematical model (HICKS et al., 1987) is presented. It only addresses the case of atmospheric pollutants that are unidirectionally transferred to the surface.

Acid deposition occurs though several possible pathways, dry deposition, wet deposition (deposition though precipitation) and occult deposition (deposition of cloud droplets). The latter two shall not be discussed in this study. Dry deposition is the direct transfer of pollutants from the atmosphere to the earth's surface or the vegetation, where it is absorbed within the canopy (MONTEITH and UNSWORTH, 1990, pp. 257). The deposition may actually take place on wet surfaces such as oceans, but the word "dry" distinguishes it from wet removal of atmospheric pollutants which have been dissolved in precipitation (Brimblecombe, 1986, pp. 32). The removal of pollutants from the air by dry deposition is frequently described in terms of the deposition velocity. It links dry deposition to the atmospheric concentration of the pollutant (see equation 1).

```
F = C * Vg (1)
```

F is the flux to the surface (in g m-2 s-1), C is the concentration of the pollutant in the air (in g m-3), and Vg is the deposition velocity (in m s-1). Vg is defined as "the ratio of the rate of deposition per unit area to the concentration at some convenient standard height above ground" (Fisher, 1993, pp. 1865).

The resistance of plants against the absorption of pollutant particles is commonly described as the inverse deposition velocity (see equation 2).

```
R = 1 / Vg (2)
```

where R is the total resistance (in s m-1). It consists of three major components:

```
R = Ra + Rb + Rc (3)
```

- the resistance to transfer of SO2 through the air (atmospheric resistance Ra),
- the quasi-laminar boundary layer resistance (Rb) in the immediate vicinity of the receptor surface (molecular diffusivity), and
- the surface or canopy resistance (Rc) of the plant (including the resistance by the cuticle of the leaf and the resistance to transfer through the stomata).

Whereas Ra and Rb are dependent on wind speed and some parameters of the vegetation cover such as canopy height, Rc also depends on relative humidity/precipitation, solar radiation and temperature (HICKS et al., 1987, MONTEITH and UNSWORTH, 1990).

These three resistance parameters can be inserted into equation (1) to represent the deposition velocity:

```
F = C * 1 / (Ra + Rb + Rc) (4)
```

HICKS et al. (1987) recommend a temporal resolution of an hour for the deposition velocity estimation, since many of the factors involved, especially the meteorological parameters, change in short terms.

---

4 Format, temporal and spatial resolution of the data available for the study site

The deposition of SO2 to the ground varies with weather conditions, the state of the vegetation cover on the surface and its seasonal changes, as well as with the SO2 concentrations in the air (SMALL and SAMSON, 1989). It is therefore important to be aware of the implications of different time scales and spatial resolutions in the data fed into the model to calculate the annual deposition of SO2 to the ground:

- The study site (7km by 8km) at Halladale in Northern Scotland (see Fig. 2) comprises a typical uplands area with moorland and spruce afforestation (482ha) between 129m and 379m altitude above sea level. Vegetation attributes (leaf area index, canopy height, etc.) for the two vegetation classes in the study area were assumed to be different for summer (April - September) and winter (October - March) [vector data].
- Altitude above sea level: digital elevation model (DEM) at 100m resolution [raster data]
- The atmospheric SO2 concentrations were available as monthly mean values [point data].
- An automatic weather station (AWS) provided data on air temperature, wind speed, relative humidity, solar radiation and rainfall were available hourly for a whole year (18-03-1993 to 17-3-1994) [point data]. Since the weather data were supplied from only one weather station, some assumptions regarding the spatial variability of these parameters have to be made. It was assumed that the meteorological parameters change over space due to the influence of the terrain. Simplifying the real world situation very drastically, two meteorological parameters, temperature and wind speed, were chosen as examples to be extrapolated over the DEM, using altitude above sea level as the only factor influencing the spatial variation. Other factors such as aspect and exposure of the terrain were not taken into account, nor was the potential for e.g. temperature inversions to change the patterns for parts of the area.

For processing purposes, the raster domain was preferred to a vector representation, as raster suit the continuous nature of the data (elevation, weather data) better and the implementation of the model is computationally faster and simpler than in the vector domain. The spatial resolution of the data sets was set to a common value, the resolution of the DEM (100m), in order to make the model development in FORTRAN simpler. However, for the purposes of the GIS-only model, this step was only implemented to aid the comparison with the coupled model. In a GIS system, raster data sets can be integrated easily without the user having to worry whether the extents and resolution of the data sets are the same.

5 Hardware and software configuration

The aim of this study was to obtain a fast and accurate modelling environment in order to improve the understanding of sulphur dioxide deposition to different vegetation surfaces in a spatial context, i.e. boundary conditions varying with terrain and weather conditions. This involved the testing of many different scenarios through changing model parameters and equations. In general, the computing requirements for running environmental models are very high. Large amounts of data have to be processed and complex equations and algorithms to be calculated repeatedly over a large number of iterations.

The model was implemented on an Alpha server 2000 4/233. ARC/INFO version 6.1 (Grid Module, ARC Macro Language AML), ORACLE 7 (with PL/SQL) and FORTRAN 77 were chosen as modelling tools for the different test versions of the model implementation.

6 Model implementation in the different systems environments

The acid deposition model is set up as an iterative accumulation of SO2 over the period of one year (results in kg ha-1 year-1). The temporal resolution of the AWS data was used as the minimal time step for the iterations of the model (8760 iterations). The model itself is quite simple. It consists of one loop over all the temporal iterations, where the AWS data are read in and the deposition per time step is calculated and accumulated. The parameters in the equations depend on the state of the vegetation over the seasons, the values of the weather parameters, the DEM (with which temperature and windspeed are calculated), and the SO2 concentrations in the atmosphere. Both the "GIS-only" and the coupled systems implementation were designed to be as similar to each other as possible (see Figure 1) in order to better compare their performance. Therefore, most of the methodological considerations that were determined for the model in general, are valid for both implementations.

```
[transform ARC/INFO Grid data into ASCII files]
DO for 1 year (8760 hours)
  READ AWS and SO2 concentration data for 1 time step
  CHECK season (for selecting the appropriate vegetation attributes)
  CHECK rainfall/humidity/solar radiation (for selection of the equations)
  CALCULATE temperature and wind speed matrices/grids
  CALCULATE the SO2 deposition and ACCUMULATE over the year
  [WRITE dry deposition values into an ASCII matrix]
END
[transform the ASCII file into ARC/INFO]
```

Figure 1: Pseudo code for the implementation of the dry deposition model (lines in [ ] indicate steps added for the FORTRAN routine)

For the coupled GIS/FORTRAN implementation, the ARC/INFO raster data sets were exported into ASCII matrices before running the model. The resulting accumulated annual dry deposition per cell were then converted back to the GIS system for display and further analysis. In order to facilitate the exporting and importing of files between the GIS as a pre- and post-processor and the FORTRAN routine, the FORTRAN routines were integrated into the AML environment, so that the user did not have to go through the cumbersome process of manually converting the raster data into ASCII format, running the model outside GIS and converting the results back into ARC/INFO grid format.

In the GIS-only implementation, the accumulation of hourly deposition over the year had to be done differently from the FORTRAN approach, since it was not possible in ARC/INFO to just overwrite a grid data set with another one of the same name. To get round this problem, grids with new names are calculated at each iteration, and the old data sets are deleted (so that the used disk space at any time is minimised).

7 Preliminary results

In general, the model's accumulated deposition values were very close to estimates from SO2 deposition monitoring at the site, with considerable differences between moorland and forest areas and increasing deposition with altitude as expected (see Figure 2).

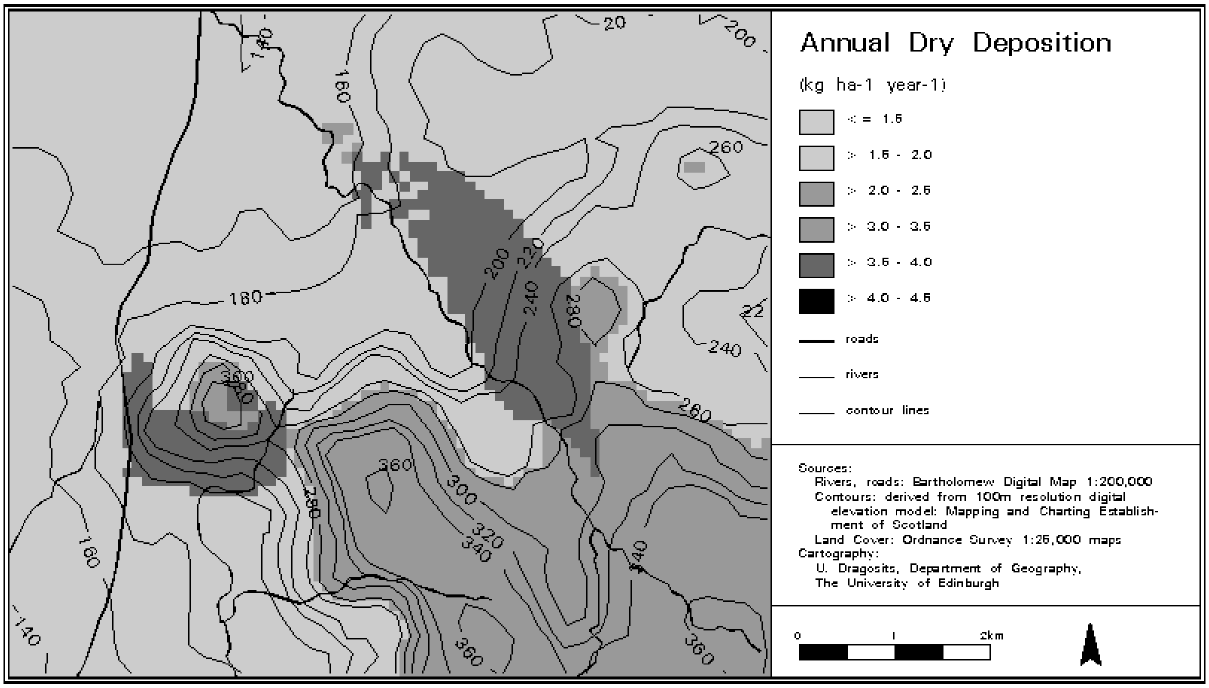


Figure 2. Accumulated SO2 dry deposition. The forest is clearly visible as the area with high deposition.

The respective advantages and disadvantages of the two approaches - GIS-only and coupled systems - can be compared:

As far as numerical accuracy is concerned, there was no difference between the two implementations.

One of the reasons why large environmental models are rarely implemented completely within GIS, but rather combine GIS and other methods or completely exclude GIS, is that GIS models are significantly slower to execute. Therefore, one of the main aims of this study was to quantify these differences in performance, i.e. measure the runtimes (CPU times) and to investigate how to optimise performance.

Whereas the whole coupled system used up only approximately 2.5 minutes of CPU time, the "identical" GIS-only implementation inside ARC/INFO Grid took nearly 21 hours, i.e. used about 500 times more computing resources to perform the same task.

As far as ease of use and model programming are concerned, there are some differences to observe, because of the features of the two different application languages. Whereas, in FORTRAN, it is necessary to build loops over arrays to reference individual cells in turn, the spatial datasets are transparently handled in the GIS/4th generation language environment. However, reading in ASCII data, creating indexed variables etc. is less straightforward.

In conclusion, the main disadvantage of the GIS-only approach is performance. Several reasons contribute to this problem:

ARC/INFO AML is an interpreted language. Each of the commands has to be translated before execution, no compiler optimisation is done in AML, as FORTRAN does. AML is based on a 1960's computer language and has not been fundamentally changed since then (HESLEY, 1994). In order to analyse the excessively long run times of the GIS implementation more closely, it was tested for performance several times during its development. For instance, just reading the AWS data line by line from an ASCII file without performing any calculations takes 2 CPU minutes, the same time as running the entire coupled implementation. Adding conditions to check the season (summer, winter) and the meteorological variables in order to determine which of the land cover data sets and which of the equations was selected, resulted in a CPU time of approx. 7 minutes. Also, during each iteration, two to three raster data sets have to be deleted - this adds up to more than 2.3 CPU hours. Creating the raster data sets themselves takes up the remaining 18.5 hours.

The main reason why it takes so long to process a single grid is that not only does the raster need to be processed, but its topology has to be built, including the generation of attribute and statistics tables. If this obligatory part of (GIS-specific) processing could be omitted between iterations, it would be expected to speed up the model considerably. Therefore, the number of raster data sets to be created over the model duration of one year is critical.

8 CPU optimisation for the GIS model

The next step was to investigate how the GIS-only model could be optimised without compromising the results. In order to achieve this goal, sensitivity testing had to be undertaken. The obvious solution for reducing the runtime is to minimise the number of grids created over the model duration. There are several ways of achieving this:

8.1 Reducing the temporal resolution

Most apparently, the temporal resolution, which determines the number of iterations to be carried out, could be reduced. However, it was uncertain how this would influence the model's accuracy. In order to test this, the AWS data were aggregated to different degrees by calculating 2-, 3- and 6-hourly average values. These different temporal resolutions, while reducing the runtime of the model to half, to a third and to a sixth respectively, proved to change the results of the deposition model significantly, as was. Whereas using 2-hourly averaged AWS data seems to be marginally justifiable in terms of the results (8% difference for moorland and 10% difference for forest with respect to the 1-hourly results), the 3- and 6-hourly results clearly misrepresent the annual deposition to a degree that is not tolerable (differences of over 50% for the 6-hourly means and about 70% for the 3-hourly means). Therefore, this was not considered an adequate means for improving performance.

8.2 Reducing the number of raster data sets created during each iteration

The other evident solution was to reduce the number of raster data sets created during each iteration. Depending on the conditional statements, either two and three grid data sets were created at each timestep. These were windspeed varying with altitude and SO2 deposition at every time step and temperature varying with altitude for approximately one third of the iterations.

In order to reduce this overhead, spatial temperature and wind speed data sets were calculated in advance outside the model, which were then accessed from inside the model. After determining the range of AWS values for windspeed and temperature over the model duration, grids were created for each integer value of the temperature and wind speed range, i.e. each full degree Celsius of temperature and each metre per second of wind speed. These precalculated rasters were then used in the model as look-up data sets. Once these ready-to-use temperature and windspeed surfaces were created, the model was almost twice as fast (just over 11 hours). It was uncertain how this simplification using pre-produced spatial data sets for each full degree Celsius of temperature and each m s-1 of wind speed would affect the deposition estimates. However, the differences in the attributed to this simplification were minimal (0.5% for moorland and 0.3% for forest). Therefore, this simplification can be justified to speed up the performance of the GIS implementation.

8.3 Creating "virtual grids"

ARC/INFO Grid has a specific structure, the "DOCELL loop" (ESRI 1991 a,b), which allows the user to process "temporary grids", i.e. no files are written to disk, the raster data are kept in RAM and discarded after exiting the loop. This sounds like an ideal solution, as the use of temporary grids would bypass the creating of topology and also the need to delete intermediate results from disk space. However, there is an internal limit set inside the DOCELL structure restricting how many of these "virtual" grids can be temporally stored in RAM or paged out to disk space. The maximum number of iterations executable inside a DOCELL loop depends on how many complex mathematical operations have to be carried out during any one iteration. It is possible to create the maximum possible number of temporary grids inside this DOCELL loop, write the intermediate results to disk space, leave the DOCELL loop discarding the "virtual grids" and then re-enter the DOCELL structure. For the model employed here, this allowed 25 iterations to be run before having to write non-temporary raster data sets to the disk. This results in a total runtime of 3.75 CPU hours, a reduction which is quite remarkable. However, this approach was not considered sound programming practice, as any increases in the complexity of the model equations would slow the model down considerably and need thorough testing for the new (lower) limit of maximum possible iterations inside the DOCELL loop. This looping structure was obviously not constructed for the purpose of running iterative models.

8.4 Modelling inside the GIS attribute database

In another attempt to speed up the model by entirely bypassing the creation of new files, the model was implemented inside ARC/INFO's database system. An AML accessing the attribute data for each grid cell directly was built, which resulted in a total CPU time of ca. 3 hours (5 hours including the calculation of temperature and windspeed surfaces at each time step). However, the more complicated the model equations get, the less suitable this approach becomes due to the limited mathematical functions inside the system's database.

8.5 Modelling inside an external database

Therefore, it was thought that employing an external database linked to the GIS such as ORACLE in combination with

PL/SQL, an application development tool extending the standard database language, would be the solution to this problem. However, because of the many necessary updates of the database and their impact on the size of the transaction log files, this attempt had to be aborted, as the DBMS was used for a purpose it was not designed for. Theoretically a total runtime of 44 CPU hours was estimated (including the calculation of temperature and windspeed at each iteration).

The solution to this problem would be to bypass ORACLE's updating the database at every time step and run the model as an embedded SQL written in FORTRAN or C. Compared to the initial GIS/FORTRAN implementation, this would be a more complex system utilising 3 software components rather than 2. On the other hand, the more complicated the model gets and the more data sets which would need explicit transformation between the GIS and ASCII matrices, the more feasible it is to take the data directly from the linked external database ORACLE and feed the results back into it after calculating the deposition outside. Furthermore, the development of ARC/ORACLE should integrate the systems even more tightly.

	CPU hours	specification
ARC/Grid+FORTRAN	00:02:30	
ARC/Grid only	20:41:00	temperature & windspeed calculate inside model
	11:09:00	pre-calculated temperature and windspeed
	3:43:00	"DOCELL loop"
ARC/Tables	4:48:00	temperature & windspeed calculate inside model
	2:53:00	pre-calculated temperature and windspeed
ARC+ORACLE+PL/SQL	44:12:00	temperature & windspeed calculated inside model

Table 2: Summary: CPU times of running the various implementations of the dry deposition model.

8.6 Other possibilities

The different implementations of the model presented in this paper comprise only a small number of possibilities. There is large scope for other variations. Only a few possibilities shall be mentioned here:

Direct access to the GIS data structure using a 3rd generation programming language in a coupled systems implementation would make the conversion of data between the GIS and the outside model redundant and create a more tightly coupled systems environment.

In this study only one GIS software, ARC/INFO, was explored. An alternative would be e.g. the public domain GIS GRASS, where users can build in their own code in C. Image processing software such as ERDAS IMAGINE could be considered, as it provides fast raster processing capabilities. Other possibilities are "visualisation tools" such as PV-WAVE, which might be considered as alternative modelling environments.

---

9 Discussion and evaluation

The project provided results from the experiments described in the above summary and additionally gave more general insights into the usefulness of GIS for environmental modelling. The results have given the opportunity to assess the advantages and disadvantages of the chosen approaches to implementing an acid deposition model.

In comparing the two approaches of GIS-only and integrated/coupled systems inplementations of the SO2 deposition model generally, there is no absolute criterion to prove that either of them is "better" than the other. Depending on the purpose of the modelling study, there are reasons why one or the other implementation is more suited to the specific aims.

If the main purpose of an environmental model is to explore the data, quickly test some ideas, immediately visualise the results or develop a small prototype over a very confined study area and with few iterations, and/or the modeller does not have the necessary knowledge to program his own fast raster processing tools, simply wants the "comfort" of not having to bother with building loops over the cells of the study area and thinking about extents, resolutions etc., and does not care about the runtime of the model then a GIS-only implementation could be ideal.

However, if the modeller knows exactly what he/she wants, has to calculate many scenarios or to run the model over several years, has the necessary programming knowledge and takes care of all the details regarding data input and output, then an approach utilising the fast matrix operation capabilities of a 3rd generation language such as FORTRAN or C can prove ideal for implementing the model.

It should also be noted that a hybrid approach, such as the GIS/FORTRAN or the GIS/DBMS/Embedded SQL implementation presented here, relies heavily upon the GIS system for the provision of the spatial data, for display, and for any analysis steps required (such as finding the mean deposition for each land cover class or altitude range etc.). Also, in this study, the area modelled is only 70 by 80 raster cells. In the very likely case of using a larger study area, the shifting of the time-consuming raster processing of thousands of iterations over a large number of cells to a more appropriate tool outside the GIS seems advisable. In fact, the larger the number of cells per raster, the more essential this becomes.

---

10 Conclusions and future avenues of research

In general it can be concluded that space is an essential concept in acid deposition modelling. Since GIS is defined to employ space as its basic paradigm, it provides important additional insights to traditional views of environmental modelling. However, with current hardware and software, the processing of CPU intensive environmental models such as an iterative acid deposition model in a pure GIS implementation is a fairly slow task, even when optimised. Although a GIS-only model provides the facilities to build and execute such models, there are severe performance trade-offs compared to traditional fast matrix processing tools such as third generation programming languages, for instance FORTRAN.

On the other hand, GIS are ideally suited for the provision and preparation of spatial data sets, high quality mapping, analyses of spatial relationships such as proximity and the assessment of variability in data sets with a spatial component, e.g. topography, temperature or rainfall. Depending on the purpose of a modelling study, each approach has its advantages and disadvantages.

GIS developers have to address the problem of poor performance before these systems can be used in isolation for complex modelling tasks. Until then, it seems that the best avenue for model development would be to combine the advantages of GIS and 3rd generation languages, possibly with external databases. In this way each component is assigned the task it does best, i.e. GIS is used as a pre-processor and post-processor, but the mathematical models are executed elsewhere.

References

Aronoff, S. (1989): Geographic Information Systems. A Management Perspective. Ottawa: WDL Publications.

Brimblecombe, P. (1986) Air composition and Chemistry (Cambridge Environmental Chemistry Series). Cambridge, New York, Melbourne: Cambridge University Press. pp. 224.

Carmichael, G.R., Peters, L.K. (1984b) An Eulerian Transport/Transformation/Removal Model for SO2 and Sulfate - II. Model Calculation of SOx Transport in the Eastern United States. Atmospheric Environment, Vol. 18, No. 5, p. 953-967.

Collins, S., Smallbone, K. and Briggs, D. (1994) A GIS approach to modelling small area variations in air pollution within a complex urban environment. GIS Research UK 1994 - 2nd International Conference. p. 301-307.

ESRI (1991a) GRID Command References. ARC/INFO Command References Rev. 6.0. Redlands, California.

ESRI (1991b) Cell-based Modelling with GRID. Analysis, display and management. ARC/INFO User's Guide Rev. 6.0. Redlands, California.

Fedra, K. (1993a) GIS and Environmental Modelling. In: Goodchild, M.F., Parks, B.O., and Steyaert, L.T. [eds] (1993) Geographic Information Systems and Environmental Modelling. Oxford: Oxford University Press, p. 35-50.

Fedra, K. (1993b) Distributed Models and Embedded GIS: Strategies and Case Studies Integration. 2nd International Conference on Integrating GIS and Environmental Modelling, Breckenridge, Colorado, 1993.

Fisher, B.E.A. (1983) A Review of the Processes and Models of Long-Range Transport of Air-Pollutants. Atmospheric Environment, Vol. 17, No. 10, p. 1865-1880.

Hesley, Z.J. (1994) An evaluation of the performance of the ARC Macro Language (AML) with the aim of producing a series of guidelines on writing efficient AML applications. Unpublished Msc. Thesis, Department of Geography, University of Edinburgh.

Hicks, B.B., Baldocchi, D.D., Meyers, T.P., Hosker, R.P., Jr., Matt, D.R. (1987) A preliminary multiple resistance routine for deriving dry deposition velocities from measured quantities. Water, Air and Soil Pollution 36, p. 311-330.

Kemp, K.K. (1992) Environmental Modelling with GIS. A Strategy for Dealing with Spatial Continuity. GIS/LIS 1992 Annual Conference and Exposition, San Jose, California. p. 397-406.

Lee, T.J, and Pielke, R.A. (1993) GIS and Atmospheric Modeling: A Case Study. 2nd International Conference on Integrating

GIS and Environmental Modelling, Breckenridge, Colorado, 1993.

Monteith, J.L., Unsworth, M.H. (19902) Principles of Environmental Physics. London, New York, Melbourne, Auckland: Edward Arnold. pp. 291.

Nyerges, T.L. (1992) Coupling GIS and Spatial Analytical Models. Proceedings of the 5th International Symposium on Spatial Data Handling, 1992, Charleston, South Carolina. Vol. 2, p. 534-543.

Reyes, C., Zamora, F, Legorreta, G. (1993) SIGMA: A Geographic Information System for Atmospheric Models of the Mexico City Metropolitan Area. 2nd International Conference on Integrating GIS and Environmental Modelling, Breckenridge, Colorado, 1993.

Rinaldi, G., Cavallone, G., Stanghellini, S. and Vestrucci, P. (1993) Air Quality Assessment using Geographical Information Systems. Conference Proceedings: 4th European Conference and Exhibition on Geographical Information Systems (EGIS 1993), p. 284-293.

Small, M.J., Samson, P.J. (1989) Stochastic Simulation of meteorological Variability for Long-Range Atmospheric Transport - I. Dynamic Lagrangian Models. Atmospheric Environment, Vol. 23, No. 12, p. 2813-2824.

Steinnocher, K. and Knoflacher, H.M. (1994) Entwicklung eines Rastermodells zur Simulation raeumlicher Belastungen durch Luftschadstoffe unter Verwendung von ARC/INFO Grid. Conference Proceedings "Angewandte Geographische Informationstechnologie VI, Salzburg" (AGIT'94), p. 703-711.

Trozzi, C. and Vaccaro, R. (1993) Air Pollutants Emissions Inventory and Geographic Information Systems. Conference Proceedings: 4th European Conference and Exhibition on Geographical Information Systems (EGIS 1993), p. 47-56.

U. Dragosits and C.J. Place
Department of Geography
The University of Edinburgh
Drummond Street
Edinburgh EH8 9XP
Scotland, UK
e-mail: ud@geo.ed.ac.uk and cjp@geo.ed.ac.uk
UD tel.: (44) 0131 650 2662 and CJP tel.: (44) 0131 650 2542
fax: (44) 0131 650 2524

R.I. Smith
Institute of Terrestrial Ecology
Bush Estate
Penicuik
Midlothian EH26 0QB
Scotland, UK
e-mail: ris@ite.ac.uk
tel.: (44) 0131 445 4343
fax: (44) 0131 445 3943

John K. Horne, J. Michael Jech, and Stephen B. Brandt

Spatial Modeling of Aquatic Habitat From a Fish's Perspective

---

Patchy distributions of organisms and environmental conditions are characteristic traits of aquatic environments. Yet biological and physical measurements are often averaged over large spatial scales when used in environmental models. From a fish's perspective, local rather than spatially averaged conditions determine available habitat. We found that a horizontal cell size of 40 m preserved environmental heterogeneity in habitat quality analyses. Habitat quality was quantified using prey density, prey size, and water temperature data in bioenergetic models to estimate the potential growth of adult walleye (*Stizostedion vitreum vitreum*) in Lake Erie. Maps of potential growth rates showed that walleye growth is more sensitive to changes in water temperature than to changes in prey density. Addition of a programing language or increasing the flexibility of data formats would increase the use of commercial GIS systems in environmental modeling.

INTRODUCTION

Heterogeneity in spatial distributions of organisms is a near universal trait of aquatic environments (Hensen 1911, Hardy 1935, Hardy 1936). Efforts to explain biological patchiness have traditionally correlated organism counts with measures of physical conditions (Denman and Powell 1984, Legendre and Demers 1984, Mackas et al. 1985). From a fish's perspective, simple descriptions of physical or biological structure may not adequately characterize fish habitat. Fish growth and foraging are two examples of biological processes that are influenced by local water mass properties (e.g. temperature), prey densities, and prey sizes. The resulting distributions and growth of fish are functional responses to local habitat conditions. One way to quantify habitat quality of fish is to integrate water temperature and prey distributions in bioenergetic models of fish growth. Bioenergetic models use a mass balance approach where energy gained by consumption is partitioned into growth, metabolism, and waste (Winberg 1956, Kitchell et al. 1974, 1977). Bioenergetic models can be parameterized for any species of interest and used to map habitat quality.

Traditional bioenergetic models of fish growth depict all habitat as a homogeneous volume. Water temperatures and prey densities are arbitrarily set to average values from a large environment such as a lake. This eliminates spatial heterogeneity in model calculations and implicitly sets the spatial resolution or grain of the model equal to the range or extent of the environment being modeled. An often stated criticism of this approach is that model results do not accurately reflect biological and physical heterogeneity present in the environment (Stephens and Krebs 1986). As a supporting example, Lasker (1978) showed that survival of larval anchovy (*Engraulis mordax* Girard) depends on ephemeral patches of high prey density. A steady diet of average prey densities leads to starvation.

As an alternative to using mean prey densities and temperatures, spatially-explicit bioenergetic models incorporate heterogeneous prey distributions and physiologically-important, environmental conditions in model calculations (Brandt et al. 1992). High-resolution, continuous data from surveyed transects are placed in two-dimensional matrices. Matrix elements represent cells with sizes small enough to assume homogeneous conditions within each cell. The use of small cell sizes increases the spatial resolution of model calculations greater than the spatial range of the modeled environment. The potential growth of a specified predator is estimated in each cell using measured water temperatures, prey biomass densities, and prey sizes. This approach has been used to examine spatial patterns of planktivory in Chesapeake Bay (Luo and Brandt 1993), to map seasonal growth potentials of fish in Chesapeake Bay (Brandt and Kirsch 1993), to examine predator-prey overlap as a function of water temperature in the Great Lakes (Goyke and Brandt 1993), and to define habitat quality based on predator physiology (Mason et al. 1995).

This study uses spatial variance in fish densities to set cell size in spatially-explicit bioenergetic models of walleye growth in Lake Erie. Setting cell size in any spatial model is a trade-off between maintaining heterogeneity of the environment in model calculations, and reducing data resolution to accommodate logistic constraints imposed by sampling or data processing. When using underwater acoustics to sample fish densities, each cell of the data matrix must contain a sufficient number of isolated individuals to accurately estimate the average size of fish in each cell. Cell sizes in spatially-explicit bioenergetic models have been as small as 25 m horizontal by 0.5 m vertical (e.g. Luo and Brandt 1993, Brandt and Mason 1994). The corresponding variance in fish lengths or fish densities have not been quantified as a function of cell size. We use potential growth rate to quantify habitat quality. Variance patterns in the distribution of habitat quality are then compared to patterns of spatial variance in fish density.

METHODS

Data Acquisition

As part of a Canadian-American research program, continuous measures of fish size, fish density, and water temperature were recorded at night along transects in the three basins of Lake Erie during September, 1994. Lengths of the three transects used in this study were 27.5 km in the western basin, 42 km in the central basin, and 35.5 km in the eastern basin.

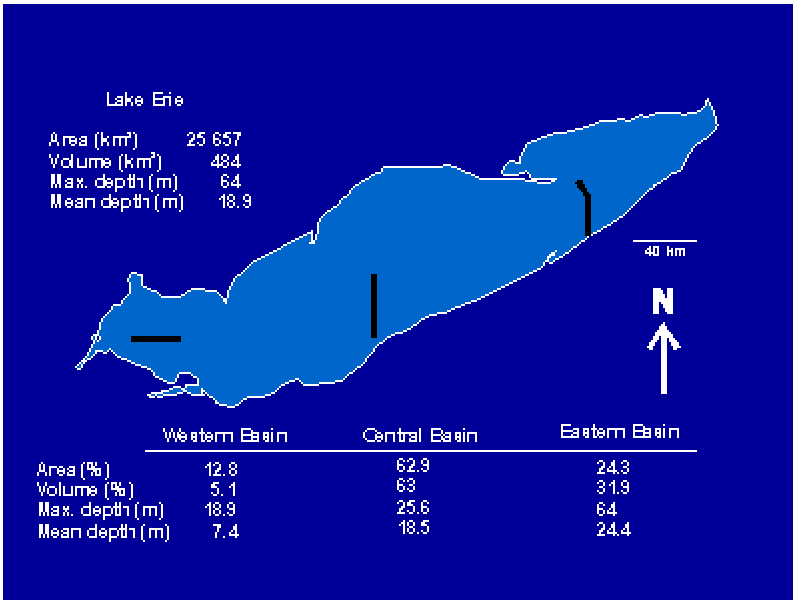


Fig. 1. Map of Lake Erie showing transects in the western (27.5 km), central (42 km), and eastern (35.5 km) basins.

Fish densities and sizes were measured using a 120 kHz scientific acoustic sonar (BioSonics Model 102, 10^o and 25^o dual beam). Sonars sample the water column by sending short, repetitive, pulses of high frequency sound in a directed beam downward as the survey vessel moves across the surface of the water. Since sound travels through freshwater at approximately 1500 meters per second (m s^{-1}), the entire water column can be quickly sampled. The sampling rate was set at one sample per second. The transducer was towed near the water surface at an average speed of 2.0 m s^{-1}. Full details of the acoustic equipment can be found in Brandt et al. (1991) and Mason et al. (1995). Water temperatures were simultaneously sampled at 2 Hz using a conductivity-temperature-depth (CTD) recorder attached to a hydrodynamic V-fin that was raised and lowered at approximately 1 m s^{-1} as the boat steamed along transects (cf. Stockwell and Sprules 1995). This 'tow-yow' sample design was converted to a two-dimensional profile of the water column by linearly interpolating temperature values in cells not sampled by the CTD. A mean value was calculated for cells with multiple measures. Cell size in the temperature data was matched to the resolution of the acoustic data.

Echo-squared integration (cf. Dragesund and Olsen 1965, Clay and Medwin 1977, MacLennan and Simmonds 1992) was used to determine the relative density of fish prey in each cell. In an initial analysis to determine the appropriate cell size for bioenergetic models, fish densities were averaged in 4 meter horizontal by 0.5 meter vertical cells. To exclude surface noise or bottom echoes, the top 2.5 m and bottom 1.0 m of the water column were not included in integration analyses. Target strengths of individual fish that were detected in each cell were estimated using a BioSonics Model 281 dual-beam signal processor (Traynor and Ehrenberg 1979, Burczynski and Johnson 1986). Target strengths of individual fish are later converted to fish lengths using an empirical target strength-length relationship (see reviews by Cushing 1973, Midttun 1984, Foote 1991). To increase accuracy of target strength estimates, only fish

o

within a maximum of 4   from the transducer acoustic axis were used in dual-beam analyses (Burczynski and Johnson 1986, Burczynski et al. 1987).

The univariate spectral analysis program BMDP1T (Dixon 1983) was used to quantify the variance in fish densities as a function of horizontal spatial scale (Jenkins and Watts 1969, Koopmans 1974, Chatfield 1984). This technique simultaneously estimates the variance of a continuously recorded variable over a range of wave numbers or frequencies. The inverse of a wave number or frequency is equivalent to the spatial scale. The range of scales examined in any analysis extends from twice the sample resolution to half the length of the data series. Two-dimensional fish density and walleye potential growth rate data were vertically-integrated over the entire water column to produce a one-dimensional data series for spectral analyses. This summation is analogous to examining fish distribution in a transparent water column from a vantage point above the water surface. The resulting spectral density estimates, a measure of variance in the frequency domain, were plotted as a function of frequency. The total variance in any data series is approximated by the area under the spectral density curve (Denman 1975). An averaging window of 0.01 cycles m^{-1} was used to compute final spectral density estimates. To permit direct comparison of spatial variance among the three transects, all spectral density estimates were centered (i.e. subtract the mean) and standardized (i.e. divide by the variance of the original series) (Denman 1975).

Model Structure

Spatially-explicit bioenergetic models of fish growth rate potential are fully described in Brandt et al. (1992). In brief, the first step in this process combines echo integration with target-strength analyses to form two-dimensional matrices of fish size and density. Echo integration analysis provides measures of relative fish densities. Absolute density is estimated by dividing relative fish density by the average target strength in each cell. Cells that lack target strength measures were assigned an average target strength randomly chosen from a Gaussian probability distribution of target strengths based on the mean and standard deviation from the entire transect. Target strengths were converted to fish lengths using Foote's (1987) empirical regression equation for clupeid fish. Prey fish lengths were converted to biomass using length-weight equations derived from samples collected in Lake Erie during September 1994 by the Ohio Department of Natural Resources.

Foraging by the predator was based on the encounter-rate model of Gerritsen and Strickler (1977). Prey density and size was used to estimate prey availability in each cell. Encounter of predators with prey was modeled as a function of reactive distance (i.e. maximum distance where predators react to prey), the ratio of predator to prey swimming speeds, and prey densities (Gerritsen and Strickler 1977, Gibson and Ezzi 1990, Persson and Greenberg 1990). Reactive distance of predators to prey was assumed equal to predator body length. Predator swimming speed was modeled as a function of predator weight (Ware 1978). Prey swimming speed was assumed negligible relative to that of the predator. Consumption of prey in each cell was estimated as the predator-prey encounter rate discounted by an assumed probability of consumption. The foraging efficiency of walleye was assumed to be 0.001. This figure represents a 10% efficiency in prey detection, a 10% predator attack rate, and a 10% rate of prey capture and ingestion. The number of prey consumed was converted to biomass consumed in each cell using length-weight relationships for each prey species. Consumption by a predator is limited by the amount of prey that can be consumed, assimilated, and evacuated over a specified time period. Temperature- and weight-dependent functions are used to limit species-specific daily consumption. Final output of the bioenergetic model is potential growth rate (g g^{-1} day^{-1}) of a specified type and size of predator in each cell.

The bioenergetic model used in this study was parameterized for a 450 mm, 990 g walleye. Predator weight was estimated using a length-weight regression based on samples of walleye taken in the western basin of Lake Erie during September 1988 (Hartman 1989). Predators were assigned an energy density of 5.0208 kJ g^{-1} wet weight (Hewett and Johnson 1992). Variable and parameter values used in consumption, respiration, specific dynamic action (i.e. basal metabolism), egestion, and excretion equations matched those used for walleye by Kitchell et al. (1977).

Walleye eat a diverse set of prey whose proportion in the diet depends on prey availability and location (Bur and Witzel 1995). Walleye primarily eat age-0 fish (Chevalier, Forney 1974, Knight et al. 1984) and select for soft-rayed prey (Forney 1974, Knight et al. 1984). In Lake Erie, gizzard shad (*Dorosoma cepedianum*) and rainbow smelt (*Osmerus mordax*) are two dominant components of walleye diet (Bur and Witzel 1995). Based on data from stomach samples, a 450 mm walleye consumes prey ranging from 40 mm to 135 mm in length (cf. Fig. 6, Knight et al. 1984). When calculating the consumption of prey biomass, all fish within this length range were assumed to be shad in the western basin, an equal mixture of shad and smelt in the central basin, and smelt in the eastern basin (cf. Bur and Witzel 1995). Gizzard shad and rainbow smelt were assigned energy densities of 4.1236 kJ g^{-1} (Pierce et al. 1980) and 5.4392 kJ g^{-1} (Foltz and Norden 1977) wet weight, respectively. The smelt energy density value used in this study was not significantly different from the one derived by Rand et al. (1994) for smelt sampled during September in Lake Ontario.

RESULTS

Patterns of spatial variance in fish densities were consistent among the three Lake Erie transects (Fig. 2). Maximum spatial variance of relative fish density occurred at the largest scale (360 m) that significantly contributed to the spatial variance in each transect. Spatial variance in fish density decreased from large to intermediate spatial scales, plateaued at intermediate scales (approximately 180 m), and then decreased at varying rates to the smallest scale analyzed (8 m). The rate of decrease in spatial variance from intermediate to small scales was greatest in the western basin.

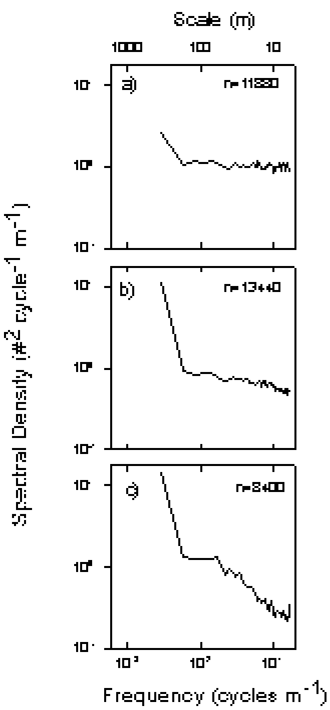


Fig. 2. Spectral density estimates of relative fish densities in a) western, b) central, and c) eastern basins of Lake Erie plotted as a function of frequency (bandwidth 0.01, centered, normalized). Periods (m) are shown on upper X axis.

Concentrations of spatial variance, characterized by distinct peaks in spectral density plots, were not observed in any of the three transects. Peaks in spatial variance are biologically interpreted as characteristic patch sizes. The rapid drop in spatial variance over a short range of scales indicates that fish densities were most heterogeneous at the largest scales sampled. In the eastern and central basins, spatial variance in the horizontal distribution of relative fish density did not dramatically decrease at scales smaller than 120 m. Variability in fish densities monotonically decreased at scales smaller than 120 m in the western basin. A cell size of 40 m horizontal by 0.5 m vertical was chosen to calculate average temperature, absolute prey size and density, and predator potential growth rates. Increasing the horizontal resolution above 40 m (i.e. smaller cell sizes) would not significantly increase the amount of spatial heterogeneity included in bioenergetic calculations. A cell size of 40 m horizontal also increases the probability that several individual targets would be present in each cell for the conversion of target strength to fish lengths.

In general, the thermal structure of the Lake Erie water column typified a two-layer, stratified lake. A warm upper layer (~20 ^{o}C) overlaid a cooler layer when the water column was deeper than the thermocline (approximately 25 m). Maximum depths in the western and central basin transects did not exceed thermocline depth (Fig. 3). Minimum temperature of the cool layer in the eastern basin transect was 4.9 ^{o}C.

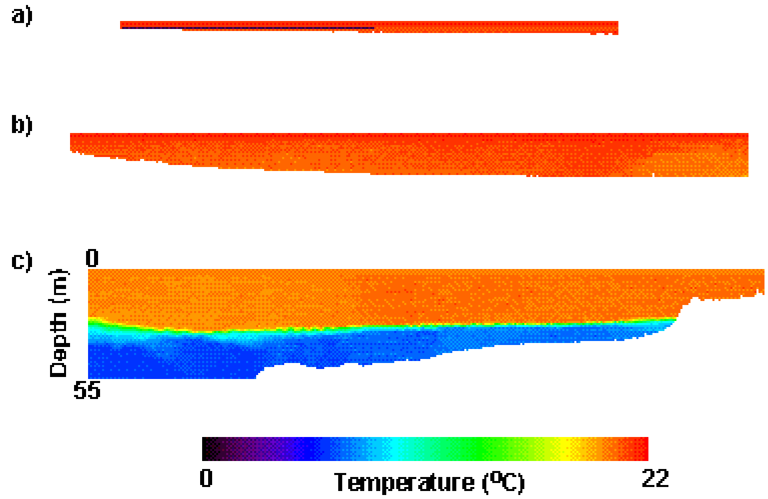


Fig. 3. Average temperature (^{o}C) of transects sampled in the a) western, b) central, and c) eastern basins of Lake Erie. Cell size is 40 m horizontal by 0.5 m vertical.

The spatial distribution of fish sizes, identified from single acoustic targets, differed among the three transects (Fig. 4). In the western basin, aggregations of larger fish were located at the inshore and offshore ends of the transect. In the central basin, larger fish were located along bottom and in two near-surface groups. These near-surface groups of large fish coincided with two aggregations of intermediate sized fish mid-way, and at the offshore end of the transect. Along the eastern basin transect, larger fish were distributed throughout the water column offshore and were concentrated near the thermocline at the nearshore end of the transect. Intermediate sized fish were aggregated just above the thermocline along the length of the eastern basin transect.

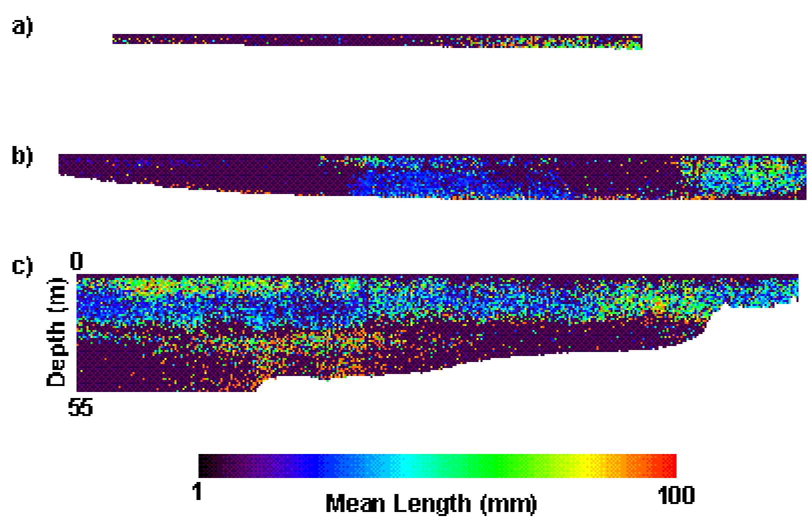


Fig. 4. Mean lengths (mm) of individual acoustic targets in the a) western, b) central, and c) eastern basins of Lake Erie. Target strengths were converted to fish lengths using Foote's (1987) equation for clupeid fish.

Patterns of variance in fish lengths (Fig. 5) closely matched patterns of average fish length along each transect (Fig. 4). High variance in fish length indicates a mixture of fish sizes within a cell. In the western basin, fish lengths varied most in the offshore third of the transect. Variance in fish lengths was highest along the bottom of the central basin transect with two vertical bands of variance in the middle and at the offshore end of the transect. In the eastern basin, fish lengths varied throughout the upper layer of the transect.

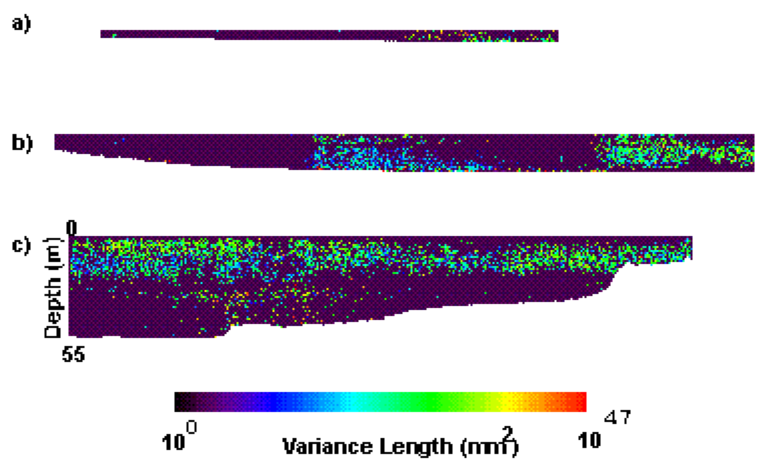


Fig. 5. Variance (mm^2) of fish lengths within cells from transects sampled in the a) western, b) central, and c) eastern basins of Lake Erie.

Densities of fish in each cell were calculated by dividing acoustic measures of relative density by the average target strength in that cell. No discernable spatial patterns of fish density were evident in the western basin transect (Fig. 6). Fish density was highly variable throughout the water column. In contrast, the central basin transect was divided into three sections by two vertical bands of dense fish concentrations. Fish densities were highest at the surface and along bottom. In the eastern basin, high densities of fish were located near-surface and along the length of the thermocline.

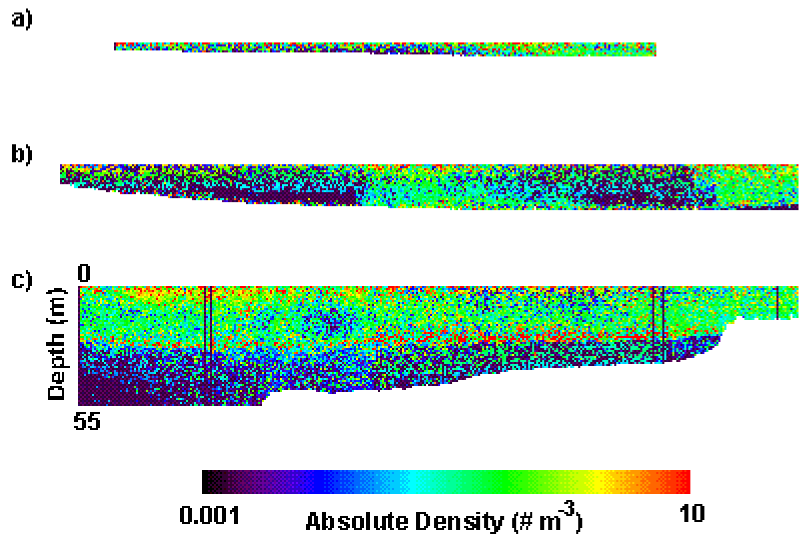


Fig. 6. Numeric density of fish (# m^{-3}) from transects sampled in the a) western, b) central, and c) eastern basins of Lake Erie.

Maps of walleye potential growth rate were based on bioenergetic calculations in each cell using the average temperature and density of suitably-sized prey (between 40 mm and 135 mm in length). Maximum potential growth of walleye in the western (Fig. 7) and central basin (Fig. 8) transects coincided with dense concentrations of prey. Temperature did not strongly influence spatial patterns of walleye potential growth in these transects due to near-homogeneous temperatures throughout the water column. Water temperature did influence patterns of walleye potential growth in the deeper eastern basin transect (Fig. 9). Areas of high potential growth matched concentrations of prey along the surface and in the water layer above the thermocline. The thin band of prey below the thermocline did not result in high walleye growth rates because of cooler water temperatures. All cells with high potential growth rates were located in cells with a minimum water temperature of 19 ^{o}C.

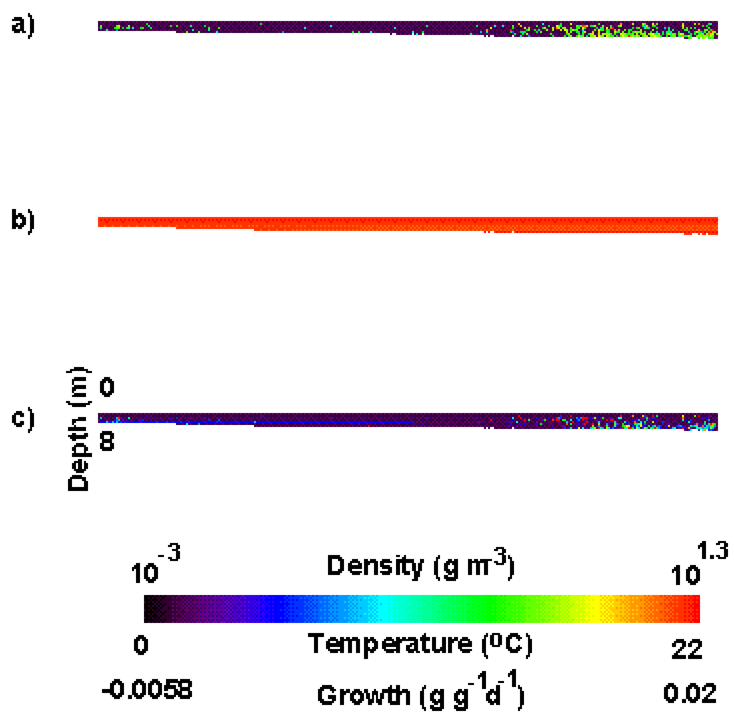


Fig. 7. a) Biomass density (g m^{-3}) of 40 to 135 mm prey fish, b) average temperature (^{o}C), and c) potential growth rate (g g^{-1} d^{-1}) of a 450 mm, 990 g walleye in the western basin of Lake Erie.

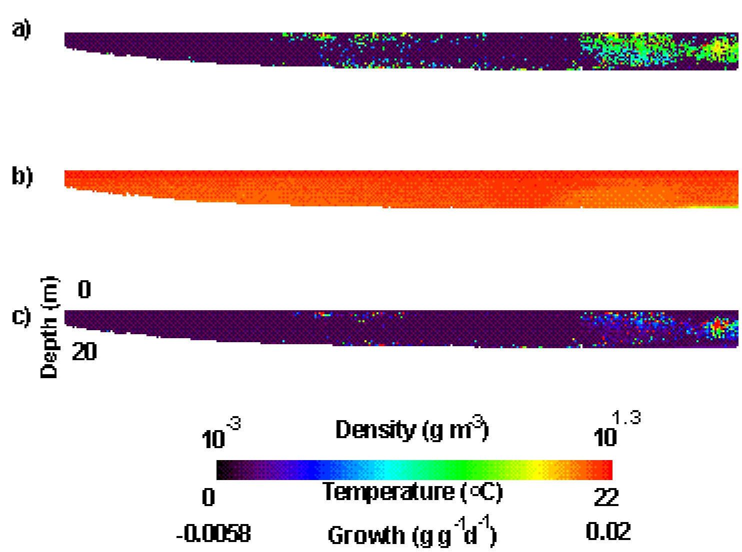


Fig. 8. a) Biomass density (g m^{-3}) of 40 to 135 mm prey fish, b) average temperature (^{o}C), and c) potential growth rate (g g^{-1} d^{-1}) of a 450 mm, 990 g walleye in the central basin of Lake Erie.

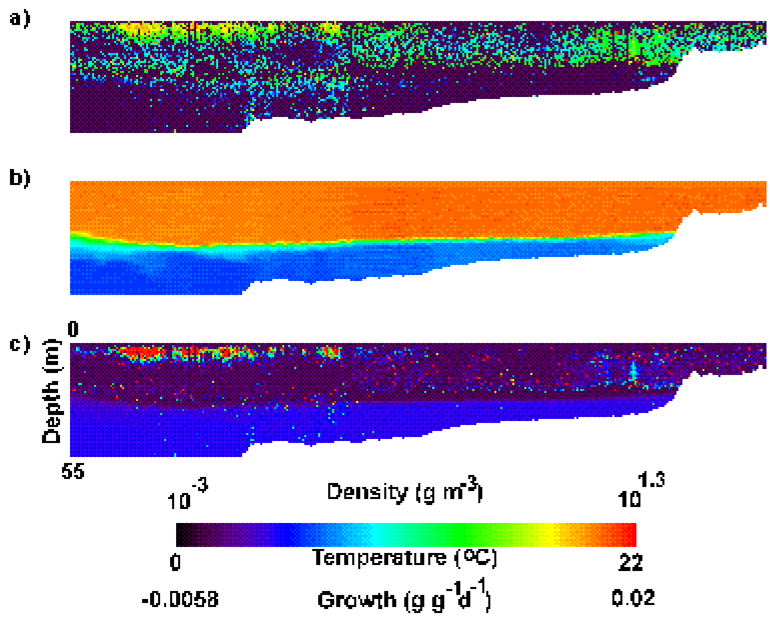


Fig. 9. a) Biomass density (g m^{-3}) of 40 to 135 mm prey fish, b) average temperature (^{o}C), and c) potential growth rate (g g^{-1} d^{-1}) of a 450 mm, 990 g walleye in the eastern basin of Lake Erie.

To examine scale-dependency in the spatial distribution of walleye habitat, spectral densities were computed for each transect using vertically-integrated potential growth rates (Fig. 10). Spatial variance in walleye potential growth rate was constant at spatial scales greater than 800 meters in all transects. At these large scales, spatial variance in walleye growth was an order of magnitude lower in the western basin than in the central or eastern basin. At intermediate scales (180 to 800 m), the large drop in spatial variance of potential growth rate forms a transition between two domains of spatial variance in walleye habitat quality. Slopes in transition regions matched those in spectral density plots of relative fish densities (Fig. 2).

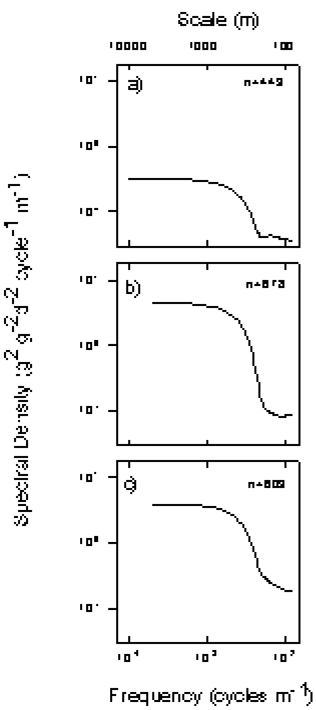


Fig. 10. Spectral density estimates of walleye potential growth rate in a) western, b) central, and c) eastern basins of Lake Erie plotted as a function of frequency (bandwidth 0.01, centered, normalized). Periods (m) are shown on upper X axis.

DISCUSSION

Growth rate potential was used to spatially index habitat quality from a fish's perspective. The technique requires high-resolution temperature and prey density data. The resulting maps of potential growth rate depict a predator's physiological response to a heterogeneous physical and biological environment. Since the entire water column is indexed, calculations are not dependent on predator distributions. Habitat quality can easily be compared among ecosystems. Summation of potential growth rates over entire water bodies can then be used as an index of fish production, to estimate fish stocking levels, to evaluate potential species introductions, and to assess the effects of environmental or anthropogenic perturbations on fish growth.

Spectral density plots were used to evaluate the distribution of spatial variance in prey densities and to choose the cell size for calculating walleye potential growth rates. Since observed patterns of spatial variance are dependent on the scale of measurement (reviewed in Horne and Schneider 1995), the choice of model resolution potentially influences interpretation of model results. If variance of a biological quantity such as density is dependent on the scale of measurement, then precision of parameter estimates also depends on measurement scale. Precision of parameter estimates can be maximized by restricting sampling and subsequent model resolution to domains of homogeneous spatial variance

(Wiens 1989). A domain of spatial variance is a range of scales with a constant or near constant slope in spectral density plots. Spatial variance domains can be demarcated using fractal geometry (Mandlebrot 1982, Sugihara and May 1990). Large changes in fractal dimension mark boundaries and indicate scales where there may be a shift in processes that generate variance in the quantity of interest (Mandlebrot 1982). The smallest homogeneous domain of spatial variance in prey distribution occurred in the western basin (40 m to 230 m, see Fig. 10). Spatial variance domains were larger in the central and eastern basins. Setting model resolution (i.e. cell size) to 40 m maintained spatial heterogeneity of prey densities observed along transects in calculations of walleye potential growth rates.

Scale-dependent spatial variance in the vertical distribution of prey density was not explicitly examined in this study. A one-dimensional spectral analysis was used to examine the vertically integrated distribution of spatial variance in prey densities. This implicitly assumes that a predator is not restricted from foraging at any depth along transects surveyed in Lake Erie. A two-dimensional spectral analysis was not computed because a square data matrix is used by the algorithm (Ripley 1981). The disproportionate length to depth ratio of survey transects would require a large number zeros to be added to 'square' a data matrix. The addition of zeros lowers the overall mean density and resulting growth rate. The partitioning of spatial variance by spectral analysis is sensitive to low means (Fasham 1978) and to the presence of zeros in count data (Horne 1995).

Small-scale spatial heterogeneity present in maps of walleye potential growth rate did not dominate the corresponding spectral density plots of walleye potential growth rate. Reduced spatial variance at small scales was a combined result of horizontal water temperature layers and the distribution of prey biomass concentrations. At any depth along the three Lake Erie transects, water temperatures were approximately 20 ^{o}C throughout the western, central, and in the upper layer of the eastern basin. Given the near-constant water temperatures, variance in walleye potential growth rate was proportional to variance in prey biomass density. Variance in potential growth rate was low at small spatial scales since adjacent cells are likely to have similar prey biomass concentrations and water temperatures. As the distance between high concentrations of prey biomass increased, spatial variance in walleye potential growth rates also increased.

Despite differences in depth and thermal structure, the quality of walleye habitat, as defined by potential growth rate, was similar among the three Lake Erie transects. In September the upper 25 m of the water column was close to the optimum temperature of 22 ^{o}C for walleye growth (Kelso 1972, Kitchell et al. 1977). In the western and central basin transects, prey were concentrated in the warmest water available. The highest concentrations of prey were also found in the warmest waters of the eastern basin transect. Comparison of walleye growth rate potentials in the eastern basin transect supports the proposal that growth rate potentials are more sensitive to changes in water temperature than to changes in prey density (Brandt 1993, Brandt and Kirsch 1993). In the isothermal water above the thermocline, growth rate potentials did not increase above a prey density threshold of approximately 15 g m^{-3}. Small differences in prey density did not influence growth rate potentials. When comparing potential growth rates at different temperatures, walleye growth rate potentials were higher above the thermocline than below for similar biomass densities of prey.

CONCLUSIONS

A spatially-explicit approach maintains the spatial heterogeneity of important biological and physical variables in environmental models. In aquatic ecosystems, bioenergetic models are dependent on remote sensing techniques (e.g. acoustic transects, CTD recorders) to provide high-resolution prey density and temperature data. Continuous prey distribution data that extend over three orders of spatial magnitude, were used to examine spatial variance in fish density distribution and to set cell size for the bioenergetic model. Data from discrete samples using traditional techniques (e.g. net samples, bathythermograph casts) potentially mask small-scale spatial variability in both biological and physical variables (Legendre and Demers 1984).

Data from transects surveyed in Lake Erie are two dimensional 'snapshots' of a small portion of a large aquatic environment. Prey and temperature distributions remain static during model calculations. The next step in the evaluation of habitat quality is to include temporally-indexed biological and physical data in the calculation of potential growth rates. The dynamics of fish growth may be examined by evaluating a series of spatial 'snapshots' at a specified temporal resolution (e.g. seasonally, dielly, hourly). A second approach would incorporate dynamic spatial models (e.g. Sklar and Costanza 1991) in the calculation of potential growth rates. Dynamic spatial models allow conditions in each cell to vary over time as both predators and prey move among cells. Individual based models (e.g. Rose and Cowan 1993) which track the spatial and temporal trajectories of predators can then be used to examine the effect of individual differences in fish size and behavior on predator-prey interactions.

Analytic and visualization techniques used in this study are not part of a commercial geographic information system (GIS), but parallel the approach advocated by GIS developers and researchers (e.g. Goodchild et al. 1993, 1996). For this study, a GIS package could be used to manage data files and display model results. However, a programming language capable of translating acoustic data to fish densities, estimating consumption based on an encounter rate model, and using bioenergetic models to calculate walleye potential growth rate is not available within current GIS versions. The increasingly complex analytic requirements of environmental models could be accommodated within a GIS by including a programming language, or by increasing the flexibility of data formats used to import or export raw data and model results. Increasing the flexibility of data formats is preferred as this modification would not limit users to a single programming language, nor restrict the diversity of environments that may be modeled using GIS packages.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation (OCE-9415740, DEB-9412113), the New York Sea Grant Institute (R/FTD-4), and by the Canadian Natural Sciences and Engineering Council(STR-0167623, SCC12100). We thank Dr. Gary Sprules for temperature data, the Ohio Department of Natural Resources for shad and smelt length-weight data, Dr. Kyle Hartman for walleye length-weight data, and Dr. Jiangang Luo for assistance in adapting the bioenergetic computer programs.

REFERENCES

Brandt, S.B. 1993. The effect of thermal fronts on fish growth: a bioenergetics evaluation of food and temperature. *Estuaries* 16: 142-159.

Brandt, S.B., and Kirsch, J. 1993. Spatially-explicit models of striped bass growth in the mid-Chesapeake Bay. *Trans. Am. Fish. Soc.* 122: 845-869.

Brandt, S.B., and Mason, D.M. 1994. Landscape approaches for assessing spatial patterns in fish foraging and growth. *in* Fresh, K. (ed.) *Theory and Application in Fish Feeding*. 390 pp.

Brandt, S.B., Mason, D.M., Patrick, E.V., Argyle, R.L., Wells, L., Unger, P., and Stewart, D.J. 1991. Acoustic measures of the abundance and size of pelagic planktivores in Lake Michigan. *Can. J. Fish. Aquat. Sci.* 48: 894-908.

Brandt, S.B., Mason, D.M., and Patrick, E.V. 1992. Spatially-explicit models of fish growth rate. *Fisheries* 17: 23-35.

Bur, M., and Witzel, L. 1995. Predator diets: Walleye in Report of the Lake Erie forage task group. *Great Lakes Fishery Commission*. pp. 6-7.

Burczynski, J.J., and Johnson, R.L. 1986. Application of dual-beam acoustic survey techniques to limnetic populations of juvenile sockeye salmon (*Oncorhynchus nerka*). *Can. J. Fish. Aquat. Sci.* 43: 1776-1788.

Burczynski, J.J., Michaletz, P.H., and Marrone, G.M. 1987. Hydroacoustic assessment of the abundance and distribution of rainbow smelt in Lake Oahe. *N. Am. J. Fish. Manage.* 7: 106-116.

Chatfield, C. 1984. *The Analysis of Time Series: an introduction*. Third edition. Chapman and Hall, London.

Chevalier, J.R. 1973. Cannibalism as a factor in first year survival of walleye in Oneida Lake. *Trans. Am. Fish. Soc.* 102: 739-744.

Clay, C.S., and Medwin, H. 1977. *Acoustical Oceanography: principles and applications*. J. Wiley & Sons, N.Y.

Cushing, D.H. 1973. *The detection of fish*. Pergamon Press, N.Y.

Denman, K.L. 1975. Spectral analysis: A summary of the theory and techniques. *Fisheries and Marine Service Research Development Technical Report* 539. Department of Fisheries and Oceans, Ottawa.

Denman, K.L., and Powell, T.M. 1984. Effects of physical processes on planktonic ecosystems in the coastal ocean. *Oceanogr. Mar. Biol. Annu. Rev.* 22: 125-168.

Dixon, W.J. 1983. *BMDP statistical software*. University of California Press, Berkeley.

Dragesund, O., and Olsen, S. 1965. On the possibility of estimating year-class strength by measuring echo-abundance of 0-group fish. *FiskDir. Skr. Ser. Havunders.* 13: 47-75.

Fasham, M.J. 1978. The application of some stochastic processes to the study of plankton patchiness. *in* Steele, J.H. (ed.) *Spatial pattern in plankton communities*. NATO Conference Series, Marine Sciences Vol. 3, Plenum Press, New York. pp. 131-156

Foltz, J.W., and Norden, C.R. 1977. Seasonal changes in food consumption and energy content of smelt (*Osmerus mordax*) in Lake Michigan. *Trans. Am. Fish. Soc.* 106: 230-234.

Foote, K.G. 1987. Fish target strengths for use in echo integrator surveys. *J. Acoust. Soc. Am.* 82: 981-987.

Foote, K.G. 1991. Summary of methods for determining fish target strength at ultrasonic frequencies. *J. Cons. Int. Explor. Mer.* 48: 211-217.

Forney, J.L. 1974. Interactions between yellow perch abundance, walleye predation, and survival of alternate prey in Oneida Lake, New York. *New York Fish and Game Journal* 27: 105-141.

Gerritsen, J., and Strickler, J.R. 1977. Encounter probabilities and community structure in zooplankton: a mathematical model. *J. Fish. Res. Bd. Can.* 34: 73-82.

Gibson, R.N., and Ezzi, I.A. 1990. Relative importance of prey size and concentration in determining the feeding behavior of the herring *Clupea harengus*. *Mar. Biol.* 107: 357-362.

Goodchild, M.F., Parks, B.O., and Steyaert, L.T., eds. 1993. *Environmental Modeling with GIS*. New York: Oxford University Press.

Goodchild, M.F., Steyaert, L.T., Parks, B.O., Crane, M.P., Johnston, C.A., Maidment, D.R., and Glendinning, S., eds. 1996. *GIS and Environmental Modeling: Progress and Research Issues*. Fort Collins: GIS World, Inc.

Goyke, A., and Brandt, S.B. 1993. Spatial models of salmonid growth rates in Lake Ontario. *Trans. Am. Fish. Soc.* 122: 870-883.

Hardy, A.C. 1935. A further example of the patchiness of plankton distribution. *Papers in Marine Biology Oceanography Deep Sea Research* 3(Suppl.): 7-11.

Hardy, A.C. 1936. Observations on the uneven distribution of oceanic plankton. *Discovery Report* 11: 513-538.

Hartman, K.J. 1989. Western Lake Erie walleye: predation, prey utilization, and the relationship with somatic growth. Master's thesis. The Ohio State University, Columbus.

Hensen, V. 1911. Das leben im ozean nach zahlungen seiner Bewohner. Ubersicht und resultaten der quantitativen untersuchungen. *Ergebn. Plankton Expdn. der Humboldt Stiftung* V.

Hewett, S.W., and Johnson, B.L. 1992. *Fish bioenergetics: model 2*. University of Wisconsin Sea Grant Institute.

Horne, J.K. 1995. Spatial variance of mobile aquatic organisms: capelin and cod in coastal Newfoundland waters. Ph.D. thesis. Memorial Univ. of Newfoundland.

Horne, J.K., and Schneider, D.C. 1995. Spatial variance in ecology. *Oikos* 74: 18-26.

Jenkins, G.M., and Watts, D.G. 1968. *Spectral analysis and its applications*. Holden-Day, San Francisco.

Kelso, J.R.M. 1972. Conversion, maintenance and assimilation for walleye, *Stizostedion vitreum vitreum*, as affected by size, diet and temperature. *J. Fish. Res. Board Can.* 29: 1181-1192.

Kitchell, J.F., Koonce, J.F., O'Neill, R.V., Shugart, H.H., Magnuson, J.J., and Booth, R.S. 1974. Model of fish biomass dynamics. *Trans. Amer. Fish. Soc.* 103: 786-798.

Kitchell, J.F., Stewart, D.J., and Weininger, D. 1977. Applications of a bioenergetics model to yellow perch (*Perca flavescens*) and walleye (*Stizostedion vitreum vitreum*). *J. Fish. Res. Bd. Can.* 34: 1922-1935.

Knight, R.L., Margraf, F.J., and Carline, R.F. 1984. Piscivory by walleyes and yellow perch in western Lake Erie. *Trans. Am. Fish. Soc.* 113: 677-693.

Koopmans, L.H. 1974. *The spectral analysis of time series*. Academic Press, New York.

Lasker, R. 1978. The relation between oceanographic conditions and larval anchovy food in the California Current: identification of factors contributing to recruitment failure. *Rapp. P.-v. Reun. Cons. Int. Explor. Mer.* 173: 212-230.

Legendre, L., and Demers, S. 1984. Towards dynamic biological oceanography and limnology. *Can. J. Fish. Aquat. Sci.* 41: 2-19.

Luo, J., and Brandt, S.B. 1993. Bay anchovy production and consumption in mid-Chesapeake Bay based on a bioenergetics model and acoustic measures of fish abundance. *Mar. Ecol. Prog. Ser.* 98: 223-236.

Mackas, D.L., Denman, K.L., and Abbott, M.R. 1985. Plankton patchiness: Biology in the physical vernacular. *Bull. Mar. Sci.* 37: 652-674.

MacLennan, D.N., and Simmonds, E.J. 1992. *Fisheries Acoustics*. Chapman & Hall, London.

Mandlebrot, B.B. 1982. *The fractal geometry of nature.* Freeman, San Francisco.

Mason, D.M., Goyke, A., and Brandt, S.B. 1995. A spatially explicit bioenergetics measure of habitat quality for adult salmonines: Comparison between Lakes Michigan and Ontario. *Can. J. Fish. Aquat. Sci.* 52: 1572-1583.

Midttun, L. 1984. Fish and other organisms as acoustic targets. *Rapp. P.-v. Reun. Cons. int. Explor. Mer* 184: 25-33.

Persson, L., and Greenberg, L.A. 1990. Optimal foraging and habitat shift in perch (*Perca fluviatilis*) in a resource gradient. *Ecology* 71: 1699-1713.

Pierce, R.J., Wissing, T.E., Jaworski, J.G., Givens, R.N., and Megrey, B.A. 1980. Energy storage and utilization patterns of gizzard shad in Acton Lake, Ohio. *Trans. Am. Fish. Soc.* 109: 611-616.

Rand, P.S., Lantry, B.F., O'Gorman, R., and Owens, R.W. 1994. Energy density and size of pelagic prey fishes in Lake Ontario, 1978-1990: implications for salmonine energetics. *Trans. Am. Fish. Soc.* 123: 519-534.

Ripley, B.D. 1981. *Spatial statistics*. Wiley, New York.

Rose, K.A., and Cowan, J.H. 1993. Individual-based model of young-of-the-year striped bass population dynamics. I. Model description and baseline simulations. *Trans. Am. Fish. Soc.* 122: 415-439.

Sklar, F.H., and Costanza, R. 1991. The development of dynamic spatial models for landscape ecology: a review and prognosis. *Ecological Studies Analysis and Synthesis* 82: 239-288.

Stephens, D.W., and Krebs, J.R. 1986. *Foraging Theory*. Princeton Univ. Press, Princeton, New Jersey.

Stockwell, J.D., and Sprules, W.G. 1995. Spatial and temporal patterns of zooplankton biomass in Lake Erie. *ICES J. mar. Sci.* 52: 557-564.

Sugihara, G., and May, R.M. 1990. Applications of fractals in ecology. *Trends Ecol. Evol.* 5: 79-86.

Traynor, J.J., and Ehrenberg, J.E. 1979. Evaluation of the dual beam acoustic fish target strength measurement method. *J. Fish. Res. Board Can.* 36: 1065-1071.

Ware, D.M. 1978. Bioenergetics of pelagic fish: theoretical change in swimming speed and ration with body size. *J. Fish. Res. Bd. Can.* 35: 220-228.

Wiens, J.A. 1989. Spatial scaling in ecology. *Funct. Ecol.* 3: 385-397.

Winberg, G.G. 1956. Rate of metabolism and food requirements of fishes. *Nauk Trudy Beloruskogo Gosudarstvennogo Universiteta imeni V.I. Lenina*, Minsk. (Translated from Russian by *Journal of the Fisheries Research Board of Canada Translation Series* 194).

John K. Horne, J. Michael Jech, Stephen B. Brandt
Great Lakes Center
Buffalo State College
1300 Elmwood Ave.
Buffalo, NY 14222-1095

Tel: (716) 878-4329
Fax: (717) 878-4009

Phenology models in complex terrain

Stuart B. Weiss
Andrew D. Weiss
Center for Conservation Biology
Dept. of Biological Sciences
Stanford Universioty
Stanford, CA 94305

Phenology of organismal development varies both between growing seasons and across habitats, and has a profound influence on the distribution and abundance of organisms. We have developed a series of GIS-based models that predict phenology across complex terrain in a variable climate. Physiological time scales for organisms are based on a combination of slope-specific insolation and daily air temperatures. Slope-specific direct radiation loads are calculated by the ARC/INFO macro program SOLARFLUX. Cloud patterns are derived from daily weather records. Thermal/insolation sums are calculated for each pixel, and phenological states of each species are tracked. Both deterministic and stochastic formulations are presented.

The model is applied to the Bay checkerspot butterfly and its larval hostplants and nectar sources. The butterfly lives in a grassland habitat with a Mediterranean-type climate. The probability distributions of emergence time of adult butterflies are mapped out across a complex landscape; spatio-temporal patterns of nectar resources and larval hostplants can be mapped out simultaneously. Because the model is based on radiation and temperature inputs as modified by landscape geometry, the model has potentially widespread applications in other ecological systems.

Spatial Modeling of Instream Biotic Integrity and Riparian Ecotone Conditions in the Big Darby Creek, Ohio

Leslie A. Zucker and Dale A. White
The Ohio State University, Columbus, Ohio

It has been suggested that the degradation of stream health associated with agricultural production can be mitigated by areas of functioning riparian ecotones, and that these areas act as buffers for areas farther downstream. The use of a geographic information system (GIS) is linked to statistical analysis of the extent and spatial distribution of riparian vegetation in order to quantify the relationship between riparian ecotone modification and instream biotic integrity, at a watershed scale. Widths and cover types (grassland, shrubland, forest land, cropland, and urban land) of riparian vegetation were interpreted from aerial photographs of 30 sub-basins within the Big Darby Creek watershed in eastern Ohio. Multiple linear regression analysis is used to identify relationships between an indicator of biological community health, the Index of Biotic Integrity (IBI), and riparian vegetation proportions and patterns. A method that accounts for time-decay of explanatory variables as a function of their hydrologic distance from a sampling location is used in the generation of spatially explicit water quality models. Model results provide information on the amounts, types, and patterns of riparian vegetation that are associated with high biotic integrity in areas of intensive agricultural development.

H. Resit Akcakaya

Linking GIS with Models of Ecological Risk Assessment for Endangered Species

A model that links GIS to models for viability analysis and risk assessment is applied to endangered species, including the Spotted Owl in the northwestern US, the Red-cockaded Woodpecker in Louisiana, and the California Gnatcatcher in Orange County, California. The model integrates landscape data on habitat requirements with demographic data to analyze risks of extinction, evaluate management options, and assess human impact on wildlife populations. Other applications of the model involve design of nature reserves, wildlife management, and population viability analysis. The model analyzes habitat data exported from a GIS, and identifies the patches of habitat that can support a population. The structure of these patches, including their locations, sizes and distances from each other, define the spatial structure of the metapopulation. The spatial structure is combined with demographic data and other information on the ecology of the species to complete a metapopulation model, which incorporates age or stage structure and density dependence for each population, spatial correlation and dispersal among populations, environmental and demographic stochasticity and catastrophes. The model performs a risk analysis, and runs multiple simulations, automatically changing parameters to analyze the sensitivity of risks to input data.

INTRODUCTION

One of the main tasks of conservation biologists is to evaluate the viability of endangered and threatened species under different natural conditions, and under alternative options for wildlife management, reserve design, and habitat protection plans. These evaluations usually ask questions about the predicted future abundance, risk of extinction, or chance of recovery of the species; and are addressed by population viability analysis (PVA), which is a systematic examination of interacting factors that place a population or species at risk of extinction (Gilpin and Soule 1986, Shaffer 1990). The factors that a PVA examines may be both natural and anthropogenic in origin, and their analysis often involves mathematical or computer models that predict the future changes in the abundance and distribution of the species in question, given information about its ecology and demography (Burgman et al. 1993).

Habitat loss and fragmentation are among the most common threats facing endangered species, making GIS-based evaluations an essential component of population viability analyses. Often habitat loss and fragmentation, combined with the natural heterogeneity of landscapes, forces species to exist in multiple populations inhabiting relatively isolated habitat patches. Such a collection of populations of the same species is called a metapopulation. The existence of multiple populations usually introduce complexities that make it impossible to evaluate the viability of the species based on PVAs performed on separate populations, and necessitate a metapopulation modeling approach.

This paper describes a computer program for building metapopulation models and performing GIS-based PVAs, and discusses its application to cases involving endangered species.

THE MODEL

The PVA program RAMAS/GIS (Akcakaya 1994a) is designed to link GIS-generated landscape data with a detailed metapopulation model for extinction risk assessment, viability analysis, reserve design and wildlife management. Descriptions of the first version of the program can be found in Akcakaya (1994b) and Akcakaya et al. (1995a); it was also reviewed by Kingston (1995). This section describes the second version of the program, which is being tested with applications to various endangered species (see the section on "Applications" below).

The program operates in four steps. First, landscape data are analyzed and the patch structure of the habitat is exported to a metapopulation model. Second, temporal changes in habitat characteristics are modeled. Third, a metapopulation model is built by combining spatial and demographic information. Fourth, simulations are run to estimate risks of extinction or decline, and to predict the abundance and distribution of individuals in the metapopulation. The essential aspects of these four model components are summarized below.

Landscape data

The function of the *Landscape data* component is to analyze GIS data to determine the spatial structure of the metapopulation as well as several population-specific demographic parameters that may depend on habitat quality. This component works in the 6 steps described below.

1. Imports landscape data from a GIS

RAMAS/GIS operates under MS-DOS, and can import landscape data in the form of raster maps exported in ASCII format from ARC/INFO, GRASS, IDRISI, and any GIS that supports bare grid ASCII format; it can also import maps in binary format from IDRISI. The requirements are: (i) maps must be in raster format, with the cells arranged in a square (not hexagonal) format; (ii) map values must be numerical; (iii) all maps must describe the landscape with the same precision (i.e., number of cells in both north-south and east-west directions must be the same); (iv) coverage of all maps must be identical (the corners of the rectangular area described by all maps must be the same); (v) maximum map size depends on the available memory, but cannot exceed 5000 rows and 5000 columns in version 2.0.

2. Creates a habitat map

Information in different maps are then combined to make a habitat suitability (HS) map, with a user-defined habitat suitability function. This function must be estimated outside the program. Most methods of estimating the HS function involve statistical procedures, using species occurrence or abundance at each location as the dependent variable and the habitat characteristics as the set of independent variables. The statistical procedures most commonly used are stepwise multiple regression methods (including logistic regression) and stepwise discriminant function analysis. The function is then entered in RAMAS/GIS, which creates a habitat suitability map by calculating the value of the function for each cell of the raster map.

3. Finds habitat patches

The program employs a patch-recognition algorithm to find clusters or groups of nearby cells that have HS values higher than or equal to a threshold habitat value, and labels them as patches. For this procedure, the program uses two parameters that determine how the species perceives (or reacts to) the patchiness of the habitat. These are threshold HS value and neighborhood distance. *Threshold HS* is the minimum HS value (as defined by the HS function) below which the habitat is not suitable for reproduction and/or survival (although individuals may disperse or migrate through habitat that has a lower HS than this threshold). *Neighborhood distance* is used to identify nearby cells that belong to the same patch. Suitable cells (as defined by the threshold parameter) that are separated by a distance less than or equal to the neighborhood distance are regarded to be in the same patch. For an animal species, the neighborhood distance parameter may represent the foraging distance.

4. Calculates demographic parameters

Next, the program calculates five demographic parameters for each patch. These are carrying capacity, initial abundance, maximum growth rate, relative fecundity and relative survival. The meaning of these parameters will be described below (see the section "Metapopulation model"). The program allows these five parameters to be calculated as any arbitrary function of patch characteristics, such as total habitat suitability in the patch, average habitat suitability, area as the number of cells in the patch, length of the perimeter (edge) of the patch, average values of input maps for the patch. In addition to these landscape variables, the functions for the five demographic parameters can also include any of the standard mathematical functions (such as min, max, log, exp, etc.)

5. Determines the spatial structure

In addition to the population-specific demographic parameters described above, the program calculates several parameters related to the spatial structure of the metapopulation. These include the location of patches, the distances among patches and migration (dispersal) rates, and spatial correlations based on these distances. These two metapopulation-level parameters, dispersal and correlation, will be discussed below. The rate of dispersal between two population can depend on one of three distance measures: edge to edge (minimum distance between the two patches), center to edge (distance from the center of source patch to the closest edge of the target patch), or center to center (distance between the centers of the two patches).

6. Exports the results to a metapopulation model

Finally, the program exports the patch coordinates, demographic parameters (carrying capacities, growth rates, etc.), the dispersal (migration) rates and the correlation structure to the metapopulation model.

Habitat dynamics

The second component of the program is designed to be used in modeling temporal changes in habitat. It allows the calculation of a time-series of carrying capacities for each population. It reads data files saved from the *Landscape data* subprogram, and outputs a set of data files for the *Metapopulation model* subprogram. One of these data files contains the main metapopulation model, and others contain the temporal changes in carrying capacities. The *Metapopulation model* subprogram inputs these files of carrying capacities and vital rates, and uses them to model habitat dynamics.

Metapopulation model

The main component of the program is where the spatial and demographic parameters calculated by the *Landscape data* component are combined with other ecological and demographic information about the species to develop a metapopulation model. The model may incorporate the factors and parameters discussed below.

Age structure or **Stage structure** within populations is modeled by a matrix model (Caswell 1989) that incorporates age- or stage-specific vital rates (survival rates and fecundities). Each population in the model can have a different stage matrix, and a different initial age or stage structure (initial number of individuals in each age or stage). The initial structures can be specified as the stable age or stage distributions. The population-specific stage matrix can be specified to change through time, by reading two files (one each for fecundities and survival rates) that contain the temporal change in the relative values of the vital rates.

Density dependence in population dynamics is modeled by modifying the mean values of survival rates and fecundities as a function of the population size (*N*). Density-dependent population growth may involve a simple ceiling model, logistic-like functions that describe contest- or scramble-type intraspecific competition (including Ricker and Beverton-Holt functions), Allee effects (i.e., density dependence at low population sizes), or Allee effects combined with density dependence at high population sizes. All density dependence functions are parameterized with the same set of parameters that include maximal growth rate (*Rmax*) and carrying capacity or ceiling (*K*), random variation in *K*, and temporal trend in *K* . Each population can have a different set of parameters.

Habitat change, e.g., habitat loss (as a result of human impact) or habitat increase (as a result of vegetation growth) can be modeled by specifying how the carrying capacity of each population changes through time. This can be done either with a constant rate, or as a time-series of carrying capacities saved in a disk file.

Environmental stochasticity is modeled by random fluctuations in vital rates and in carrying capacities. The random fluctuations can be normal- or lognormal-distributed, and can be correlated among populations. They are assumed to be perfectly correlated among age classes or stages within each population.

Demographic stochasticity is modeled by sampling the number of survivors from a binomial and the number of offspring from a Poisson distribution (Akcakaya 1991).

Catastrophes are rare events that can affect abundances (a proportion of all individuals die), vital rates (survival rates and fecundities are reduced after a catastrophe), or carrying capacities (which are reduced after a catastrophe). The spatial extent of catastrophes may be local or regional. The impact of catastrophes can be population-specific (some populations may be more prone to, or more affected by catastrophes than other populations), or they may be stage-specific (some stages, or even certain vital rates may be more affected by catastrophes than others).

Density dependence in migration is modeled by making the total rate of emigration or dispersal a function of the size of the population. The rate of dispersal/emigration can be specified either to increase or decrease as *N* increases. Migration rates can also be stage-specific.

Geographic configuration is specified by the coordinates of each population. The distance of populations from each other and their relative positions are utilized in modeling the effects of the two spatial factors discussed below.

Dispersal (migration) describes the movement of individuals among populations. In RAMAS/GIS, migration is modeled by specifying the proportion of individuals that move from each population to each other at every time step. These rates are input in the form of a migration matrix. In most cases, the rate of dispersal may be a function of the distance between source and target populations. RAMAS/GIS allows users to specify a function that describes the dependence of dispersal rates on distance. The matrix can filled according this to function, and can be edited to account for habitat corridors (by increasing the rate between specific pairs of populations) and for obstacles or geographic barriers to migration (by decreasing the rate). The migration rates may also be specified to be dependent on population size to allow for density-dependent migration (see above), or on the age or stage of the individuals to allow for age- or stage-specific dispersal tendencies.

Correlations among populations describe the similarity of environmental patterns experienced by each population. This factor is important in the "rescue effect" in metapopulations: when fluctuations are spread over a number of separate populations, the overall risk faced by the metapopulation is reduced. If the fluctuations in the environment are at least partially independent (uncorrelated), it will be less likely that all populations go extinct at the same time than if the fluctuations are dependent (i.e., synchronous or correlated). In uncorrelated environments, extinct populations will have a chance to be recolonized. Thus, correlation of environmental fluctuations has important effects on metapopulation persistence and viability (Gilpin 1988; Akcakaya and Ginzburg 1991; Burgman et al. 1993).

RAMAS/GIS models correlations by sampling the vital rates of each population from a normal or lognormal distribution which is correlated with the vital rates of other populations according to a correlation matrix specified by the user. This matrix can be specified as a function of the distance between populations (as closer populations are more likely to experience similar environmental patterns).

Risk assessment

In this component, metapopulation models built in RAMAS are used to simulate the dynamics and to predict the future of the metapopulation. RAMAS/GIS summarizes the results of a simulation with 16 types of output, some of which are superimposed. Most of these results are related to risk assessment and report risk analytical measures such as risk of extinction and time to extinction (Akcakaya 1992). Types of output include

* Risk of extinction (or decline to any level) any time during, or at the end of, the simulated time period;
* Probability of exceeding a range of population sizes any time during, or at the end of, the simulated time period;
* Distribution of times to extinction (or decline to a specific level), and cumulative time to extinction or decline;
* Distribution of times that the metapopulation will exceed a specified threshold, and cumulative time to increase;
* Abundance (of the metapopulation and each of its populations) through time;
* Metapopulation occupancy (number of extant populations through time);
* Local occupancy (for each population, the number of time steps that it was extant);
* Local extinction duration (for each population, the maximum number of consecutive time steps during which a population remained unoccupied during a typical replication);
* Final stage structure (for each population, histogram of the number of individuals in each stage at the end of the simulation);
* Metapopulation structure (histogram of the number of individuals in each population at a specific time step).

The model can be run several times, to analyze the sensitivity of results to input parameters by varying them automatically, to compare management options, or to assess anthropogenic impact by comparing outputs from simulations with parameters for impacted and non-impacted situations. The sensitivity analysis facility also allows the user to superimpose graphs from different simulations to make the comparisons easier.

APPLICATIONS

The second version of the program is currently being tested with applications to four endangered birds (California gnatcatcher, cactus wren, red-cockaded woodpecker, and northern spotted owl). These applications are briefly described below.

California gnatcatcher and cactus wren

The program is applied to California gnatcatcher *Polioptila californica* and cactus wren *Campylorhynchus brunneicapillus* metapopulations in central and coastal regions of Orange County, California. The California gnatcatcher (a federally-listed threatened species) and cactus wren are two of the three "target species" in the State of California's Natural Community Conservation Planning (NCCP) program. The target species are considered as surrogates for the conservation of coastal sage scrub, a plant community that has decreased substantially from its historic coverage as a result of urban and agricultural development, and has become fragmented.

The application started with a statistical analyses of observation locations, and GIS maps for elevation, slope, and vegetation (exported from ARC/Info) to characterize the habitat of the two birds. The habitat maps were then analyzed to determine the spatial structure of the metapopulation model. These data were combined with demographic parameters estimated from field work conducted by Atwood et al. (1995) to produce spatially-explicit, stage-structured, stochastic metapopulation models for the two species (Akcakaya and Atwood, in prep.). Simulations and sensitivity analyses with these models pointed out to the relative importance of temporal variation in vital rates (including the frequency and the magnitude of catastrophic declines in vital rates as was observed in the winter of 1994-95).

Red-cockaded woodpecker

The goal of this application of the program was to evaluate the impact of timber prescriptions (forest management practices) on the viability of a red-cockaded woodpecker *Picoides borealis* metapopulation in Louisiana. The habitat of the red-cockaded woodpecker was characterized based on the stand type (i.e, the dominant tree species), stand condition, and basal area of pine species. These variables were input into the program in the form of GIS maps exported from GRASS (Akcakaya et al. 1995b). Currently the habitat characterization is being extended to include other habitat variables such as distance to streams, and age of dominant trees in the stand. The habitat descriptions are used to evaluate the impact of forest management practices, such as a planned cut, on the carrying capacities of habitat patches and other parameters of the red-cockaded woodpecker populations inhabiting these patches. The model will then be run with different management options to compare their impacts in terms of the predicted risk of extinction, or rate of recovery of the metapopulation.

Northern spotted owl

This application focused on factors affecting the viability of the Northern Spotted Owl *Strix occidentalis caurina* throughout its range in the U.S. A second goal in this analysis was to incorporate two sources of variability in determining the threat the species faces. The study incorporated natural variation (resulting from temporal fluctuations in environmental factors) in the form of randomly distributed vital rates (survivals and fecundities). In addition, demographic stochasticity was modeled to describe chance variations in reproduction, survival and dispersal. These types of natural variation (environmental and demographic) were used to express the model results in probabilistic terms such as the viability of the species (for example in terms of the chance of survival or risk of extinction). Uncertainties that result from a lack of knowledge were incorporated in the form of parameter ranges, and were used to estimate upper and lower bounds on the estimated viability of the species.

Based on the habitat maps provided by the Forest Service, the program found 42 habitat patches. The size distribution of the patches was very skewed, with the 4 largest patches making up about 94% of the total area of all patches, and the seven largest making up about 96%. Because of the large differences in sizes of neighboring populations, the model results (risk of decline) were not very sensitive to the rate of inter-patch dispersal of juvenile spotted owls. The model predicted a large difference between upper and lower bounds on the viability of the northern spotted owl, based on the best-case and worst-case scenarios, which were parameter combinations that resulted in best and worst chance for survival. According to sensitivity analyses, the viability of the metapopulation was most sensitive to the set of vital rates used (the dependence of fecundities and survival rates on habitat), and also sensitive to the degree of spatial correlation among vital rates of the populations, and to the carrying capacities of the populations. In addition, metapopulation occupancy was sensitive to dispersal and Allee effects (Akcakaya, unpublished).

Other applications

The methodology used in the models, and the relative effects of factors discussed above on metapopulation dynamics have been discussed in several previous publications (Akcakaya 1991, 1992; Akcakaya and Ginzburg 1991; Burgman et al. 1993). Previous versions and precursors of the program have been applied to a wide variety of organisms. RAMAS/GIS (version 1.0) was used in assessing the effectiveness of translocating individuals as a management option for the endangered helmeted honeyeater (*Lichenostomus melanops cassidix*) in Australia (Akcakaya et al. 1995a).

The metapopulation model component was used to build a model for a land snail *Arianta arbustorum* metapopulation in northeastern Switzerland. The model was used to investigate the effect of population subdivision on the persistence of a land snail metapopulation and to analyze the interaction between spatial factors, population subdivision, and catastrophes (Akcakaya and Baur, in press).

A precursor to the metapopulation model component (RAMAS/space; Akcakaya and Ferson 1990) has been used to model metapopulation dynamics the California spotted owl (LaHaye et al. 1994), a subspecies of the northern spotted owl. Gibbs (1993) used RAMAS/space to model dynamics of spotted salamanders, bullfrogs, snapping turtles, swamp sparrows and water shrews in a mosaic of wetlands.

REFERENCES

Akcakaya, H.R. (1991) A method for simulating demographic stochasticity. *Ecological Modelling* 54:133-136.

Akcakaya, H.R. (1992) Population viability analysis and risk assessment. Pages 148-157 in *Wildlife 2001: Populations*. D.R. McCullough and R.H. Barrett, eds. Elsevier Publishers, London.

Akcakaya, H.R. (1994a) *RAMAS/GIS: Linking Landscape Data With Population Viability Analysis* (version 1.0). Applied Biomathematics, Setauket, New York.

Akcakaya, H.R. (1994b) GIS enhances endangered species conservation efforts. *GIS WORLD* Vol.7, November 1994, pp. 36-40.

Akcakaya, H.R., and Baur, B. (in press). Effects of population subdivision and catastrophes on the persistence of a land snail metapopulation. *Oecologia* .

Akcakaya, H.R., and Ginzburg, L.R. (1991) Ecological risk analysis for single and multiple populations. Pages 73-87 in: *Species Conservation: A Population-Biological Approach*. A. Seitz and V. Loeschcke (eds.) Birkhaeuser Verlag, Basel.

Akcakaya, H.R., and Ferson, S. (1990) *RAMAS/space User Manual: Spatially Structured Population Models for Conservation Biology*. Applied Biomathematics, New York.

Akcakaya, H.R., McCarthy, M.A., and Pearce, J. (1995a) Linking landscape data with population viability analysis: management options for the helmeted honeyeater *Biological Conservation* 73:169-176.

Akcakaya, H.R., Jackson, J., and Ginzburg, L.R. (1995b) Modeling the Red-cockaded Woodpecker Populations at Fort Polk. Report to U.S. Army Corps of Engineers.

Atwood, J.L., Fugagli, M.R., Reynolds, C.H., Luttrell, J.C., and Tsai, S.H. (1995) Distribution, population dynamics, and dispersal of California Gnatcatchers on the Palos Verdes Peninsula, 1993-1995. Proceedings of CalGnat'95 Symposium (in press).

Burgman, M.A., Ferson, S., and Akcakaya, H.R. (1993) *Risk Assessment in Conservation Biology*. Chapman and Hall, Population and Community Biology Series.

Caswell, H. (1989) *Matrix Population Models: Construction, Analysis, and Interpretation*. Sinauer Associates, Sunderland, Massachusetts.

Gibbs, J.P. (1993) Importance of small wetlands for the persistence of local populations of wetland associated animals. *Wetlands* 13:25-31.

Gilpin, M.E. (1988) A comment on Quinn and Hastings: extinction in subdivided habitats. *Conservation Biology* 2:290-292.

Gilpin, M.E., and Soule, M.E. (1986) Minimum viable populations: processes of species extinction. Pages 19-34 in *Conservation Biology: the science of scarcity and diversity*, M.E. Soule, ed. Sinauer Associates, Sunderland, Massachusets.

Kingston, T. (1995) Valuable modeling tool: RAMAS/GIS. *Conservation Biology* 9:966-968.

LaHaye, W.S., Gutierrez, R.J., and Akcakaya, H.R. (1994) Spotted owl metapopulation dynamics in southern California. *Journal of Animal Ecology* 63:775-785.

Shaffer, M.L. (1990) Population viability analysis. *Conservation Biology* 4: 39-40.

H. Resit Akcakaya
Applied Biomathematics
100 North Country Road
Setauket, New York 11733, USA
Tel: 516-751-4350
Fax: 516-751-3435
email: risk@life.bio.sunysb.edu

Thomas W Charnock, John Elgy, and Peter D Hedges,

Application of GIS Linked Environment Models over a Large Area

Abstract

Linking environmental models to a GIS makes it practical to use the models over a large region whilst maintaining the scale of detail that the model was designed for. This paper discusses the use of GIS and linked numerical models to evaluate the poten tial effects of groundwater abstraction upon crop production. Existing models of the groundwater and soil moisture system have been selected and linked with the GIS, crop models are currently being evaluated.

The most appropriate models for the soil moisture component of the system are 1-D soil moisture models that can simulate the response of the soil moisture profile to the falling watertable. To build up a picture of the soil moisture deficit as it devel ops through the growing season, the model is run repetitively across the region of interest.

Such a modelling arrangement leads to large data requirements and issues of calibration, particularly when the soil moisture model is linked to models describing the other environmental subsystems. During this project experience has shown that one must be pragmatic about addressing these issues.

INTRODUCTION

Environmental process level models have a scale range over which it is appropriate to apply them and there are pitfalls in using a model at a scale or resolution greater or smaller than that for which it was designed. GIS has the data handling capabili ty that makes it practical to perform modelling operations over a large area whilst still maintaining the scale of detail for which the model was designed. Furthermore GIS provides a medium for linking different environment models together.

This paper concerns the use of GIS and numerical models for Environmental Impact Assessment. The purpose of the methodology is to answer one simple question; what is the best estimate of the spatial variation of crop yield reduction due to water extrac tion from the Shropshire Groundwater Scheme? Clearly implicit is that cost of answering the question is to be less than the expected value of the crop loss itself. This assumption drives the methodology. Thus we cannot derive unnecessarily complex models from scratch nor sample ground truth at very fine resolutions. To this end a modelling system is being developed by linking several numerical models via a GIS. Whilst the strategy of overall system development and the problems encountered are discussed in Charnock *et al* (1996), this paper looks principally at the soil moisture and groundwater components of the system, in particular the data requirements.

Of prime concern is the movement of moisture within the soil profile in response to watertable drawdown. The models which best represent this are 1-D vertical models of the unsaturated zone, which adequately represent soil moisture conditions over a fe w tens of metres. As the area of potential effects is over a hundred square kilometres, it is therefore necessary to run the 1-D model at points across the landscape to simulate the spatial development of soil moisture deficit through a growing season.

Although this approach avoids the pitfall of using a model at a scale for which it was not designed other issues are raised. These include data requirements and model calibration, as well as choosing an appropriate regular or irregular spacing at which to run the soil model.

BACKGROUND

The Shropshire Ground Water Scheme is being developed in order to augment flow in the river Severn during periods of drought. The area of interest is the Tern river basin, Figure 1.

a

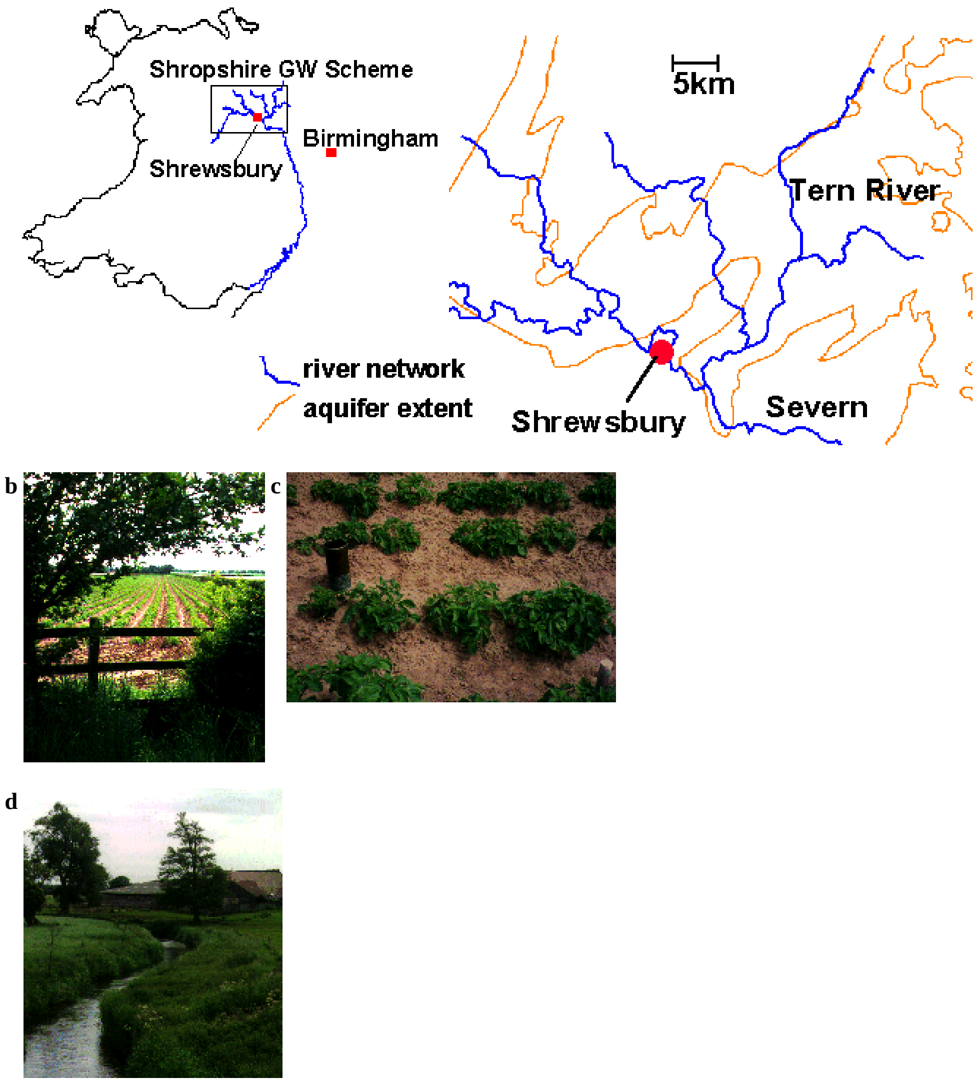


Figure 1,(a) location map, (b) field of potatoes next to Heath House Bore Hole, (c) potatoes next to Heath House Borehole with rain-gauge and soil moisture monitoring tube, (d) the River Tern with typical pastural riparion zone.

Here the watertable is close to the surface and, though there is some clay drift, for the most part the soil is in hydraulic continuity with the aquifer. Under these conditions it has been shown that pumping can potentially reduce the soil moisture avail able for crops (Hedges and Walley, 1983).

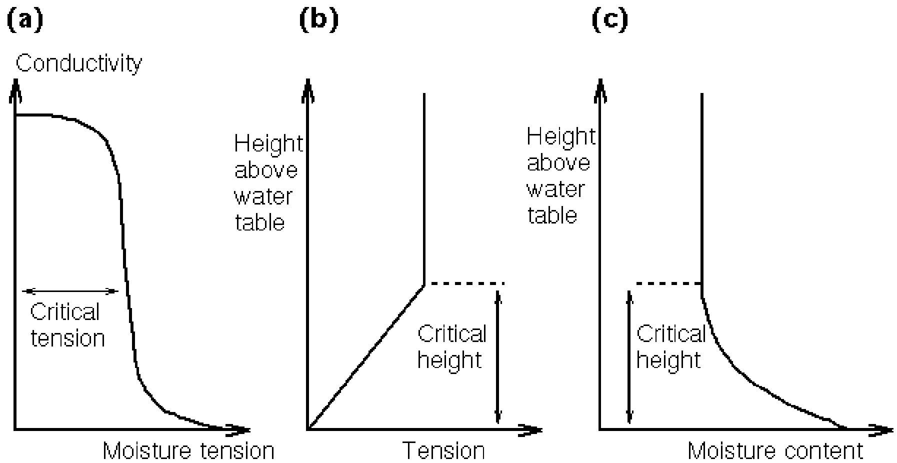


Figure 2,(a) the relationship of conductivity to tension, (b) the idealised tension profile at field capacity, (c) idealised moisture content profile.

When a soil with excess water drains it approaches a state at which tension is proportional to height above the watertable. However, a critical tension is reached at which the unsaturated conductivity drops rapidly and further drainage is negligible, Figure 2a. The soil moisture profile achieved at this tension is the field capacity profile. Hedges and Walley introduced the concept of critical height, defined as the height above the water table below which the position of the water table controls the soil moisture content (Figure 2b and c). If the watertable drops then any plants whose roots previously penetrated this zone will experience a loss of moisture. As illustrated in Figure 3.

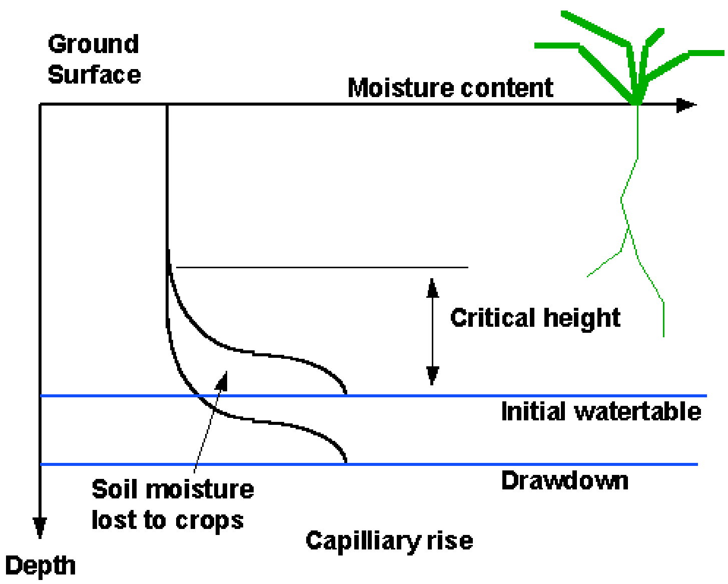


Figure 3, the loss of soil moisture due to groundwater drawdown.

Crops are vulnerable wherever the soil is in hydraulic continuity with the aquifer and the depth to the groundwater table is smaller than the critical height plus the crop rooting depth.

$D<ho+dr$ (1)
D is the depth to watertable.
ho is the critical height.
dr is the crop rooting depth.

Hedges (1989) used this simple rule within a GIS in order to delineate vulnerable areas, and used a map of predicted drawdown to identify those areas that would actually have been affected by that particular event. However, there are drawbacks to this approach; the groundwater table can be expected to drop naturally in a dry year; the soil moisture profile will be naturally depleted by crop root abstraction and, the soil moisture profile will have a complex response to watertable drawdown depending on the physical properties of the soil profile.

To improve upon the basic approach of Hedges, numerical models are being applied to simulate the groundwater movement and the response of the soil moisture and crop systems more realistically.

SYSTEM DESIGN

The overall system design and the issues raised by this approach to development are discussed in Charnock *et al* (1996), and will only be briefly reviewed here. Because of constraints of time and money the strategy of system development has been one of linking existing components together with a network of communicating programs. Minimal changes are made to existing code. It is doubtful whether any other approach would be cost effective.

GRASS is the GIS used (see Shapiro *et al*, 1993). This is a raster based system with some vector capability. The groundwater system is modelled by Modflow (McDonald and Harbaugh, 1988) which is a three dimensional finite difference formulation. T he soil moisture system is modelled with SWMS_2D (Simunek *et al*, 1994), which simulates unsaturated flow and solute movement in 1 or 2 dimensions. For this project only 1 dimensional moisture flow is required. A suitable crop model is still being s ought.

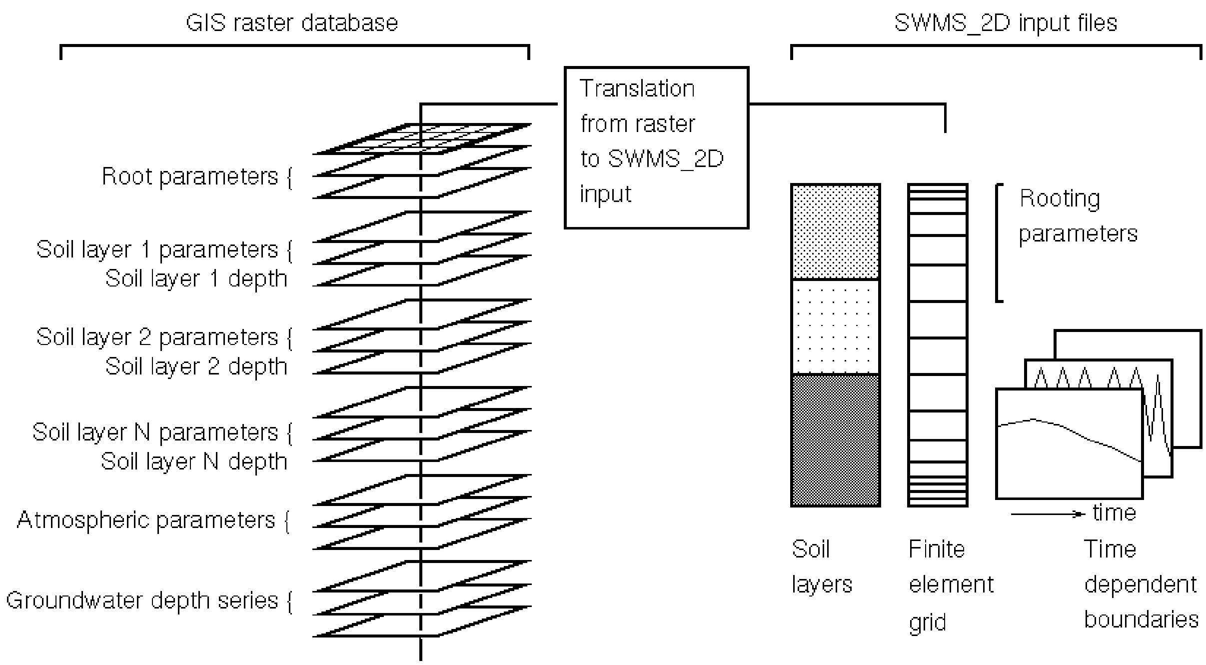


Figure 4, the link between the GRASS raster database and SWMS_2D input files.

Figure 4 illustrates the link between the SWMS_2D and the raster database of GRASS. Rasters are used to represent the spatial variation of input parameters across the landscape. The linking program samples these raster at a specified point, for which it g enerates a finite element grid of the soil column and writes the input files for SWMS_2D. A picture of the spatial variation of soil moisture deficit development is built up by calling the model successively at different points across the landscape; the a ttribute of interest is extracted from the SWMS_2D output and written into the appropriate cell of an output raster. In order to facilitate this, the link is embedded within a script which controls the calling of the model link and the build up of the out put rasters

Once any model has been linked into the GIS, channels of communication are effectively opened between the model and all the other operations and functions of the GIS, including other models. The models can thus be used in a similar manner to any other function of the GIS.

The links to the models have been designed to be implemented quickly and easily, and so they contain none of the logic of the problem to which they are going to be applied. However, once linked into the GIS they can be embedded within an appropriate ma cro language (in GRASS this is a UNIX shell script), and it is here that the logic of the problem, in terms of the subsystems to be modelled and the assumptions made about the interaction between them, are specified.

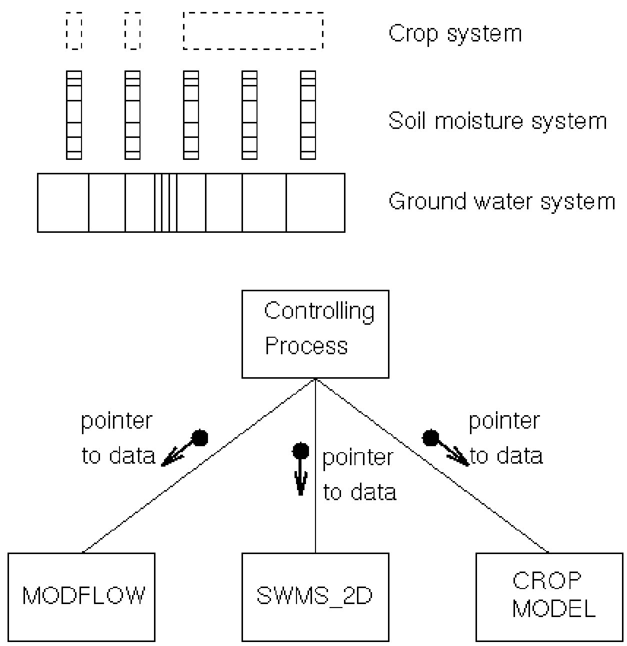


Figure 5, the modelling system viewed as component layers and as a hierarchy of control.

Figure 5 illustrates the logical arrangement of the models, which can be viewed as a layered approach. The groundwater model defines the water table as it varies across the landscape, SWMS_2D is called at an appropriate regular or irregular grid spaci ng, and the crop model will, when it is included, act either like the SWMS_2D model at points across the landscape or be applied to larger spatial units. Figure 5 also illustrates the hierarchy of control. In the simplest situation the models are run sequ entially but more complex arrangements can simulate interaction between the environmental subsystems by repetitive calls to the models (Charnock *et al*, 1996).

DATA REQUIREMENTS

As with system development, there are constraints of cost and time on the collection of data. The approach adopted has been to use secondary and surrogate sources of data wherever possible. These include geological and soil maps, standard meteorological d ata, remotely sensed data, data generated during the planning, development and operation of the Shropshire Groundwater Scheme and other documented studies.

Data requirements for the whole system include; soil properties, watertable position, potential evaporation and transpiration, precipitation, crop rooting depth and uptake parameters, aquifer transmissivity, storage and vertical conductivity, drift dis tribution and river leakage parameters.

The large scale variation of meteorological processes is an ongoing subject of research. However the system under development is intended for application to scenarios which are dry with very little precipitation, so it is reasonable to employ data from local meteorological stations utilising simple procedures such as voronoi polygons or simple interpolation.

For this project the spatial distribution of crops is required, and this will change year by year. Remotely sensed data is the most convenient way to provide timely crop information. The literature provides typical values for crop parameters, such as r ooting depth, and thus a raster of crop distribution can be classified into a distribution of rooting depths.

Aquifer properties, such as transmissivity and storage coefficient, are derived from the many pumping tests of abstraction boreholes undertaken during the history of the scheme. Watertable levels have proved to be problematical. There is an extensive n etwork of monitoring boreholes and tube wells but few have been drilled specifically for observation purposes. Thus the network is non-random, and the application of interpolation routines leads to an over estimation of watertable heights in areas where t he network is sparse.

To illustrate the problems of deriving spatially varying model input parameters from secondary and surrogate sources of data, the example of soil data will be discussed in detail.

Soil data

Soil data is needed principally for the soil moisture component. SWMS_2D uses a modification of van Genuchtens (1980) equation to describe the hydraulic properties of the soil. The moisture-suction and conductivity-suction curves are defined using inpu t parameters which include; saturated hydraulic conductivity, residual and saturated water content, as well as empirical curve fitting parameters. Evans (1990) used regression analysis to derive relationships that enable many of these parameters to be cal culated from soil properties that are easily measurable or obtainable from secondary sources such as maps.

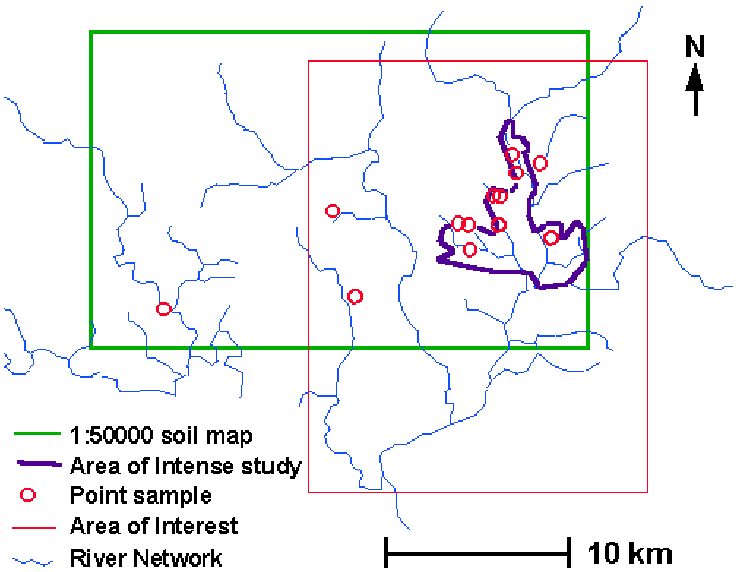


Figure 6, the availability of soil property data.

Figure 6, shows the extent of available soil data. The whole region is covered by a small scale (1:250,000) soil map, and part of the region is also covered by a larger scale (1:50000) map. The soil series is the basic mapping unit in the UK, with clas sification based on the soil profile from the surface to a depth of about 1.50m (see Soil Survey of England and Wales, 1984). The characteristic profile of a soil series is based on the particle size distribution, mineralogy, calcium carbonate content, or ganic content and the pedogenic (soil forming) processes of the soil horizons. Thus, using relationships such as those devised by Evans (1990) general purpose soil maps can be translated to maps of the spatial variation of model input parameters.

However, maps, particularly small scale maps, can be expected to have considerable inaccuracy. The types and causes of error and the importance of scale in maps generally have been well discussed elsewhere, see for example Burrough (1986). Soil maps al so have their own distinct set of limitations. The concept of sharp soil boundaries is an understandable fiction; boundaries are indistinct and within a soil series there is variation in the soil profile.

There are other soil data in addition to the two available maps. During the history of the Shropshire Groundwater Scheme several sites have been extensively studied, and there was also a very intensive survey over a small area in the Tern river valley, (see Figure 6). The point data takes a variety of forms: a detailed description of the soil profile obtained from a trial pit or augered sample; a description of soil hydraulic properties determined by direct measurement in the laboratory or field; or a record of the soil moisture regime over time. For the intensively studied area, the augered soil profile, obtained on an approximate 120m grid enabled the profiles to be classified into soil series, but with more detailed subdivisions being defined.

Crop yield varies non-linearly with respect to soil moisture parameters. Within a single soil series important parameters, such as saturated hydraulic conductivity, vary significantly. If we assume representative figures for parameters within an homoge neous unit, for example the arithmetic mean, then misleading overall yields could be produced. We must represent the naturally occurring variability within the model system. Burrough (1986) described geostatistical methods for predicting soil parameters a nd made the point that;

> " if interpolation techniques are to be of any value then they should be used if possible in conjunction with, and not instead of, conventional landscape mapping methods."

Rogowski and Wolf (1994) used a GIS to facilitate the process of combining spatially interpolated values with delineated soil map units, and produced a map;

> "that preserves the map unit boundaries but incorporates spatial variability of attributes within units".

Kiefer (1993) uses a stochastic representation of soil heterogeneity. Other authors have investigated the stochastic properties of soil heterogeneity, (e.g. Binley *et al* 1989). Rather than review their work here, one very simple stoc hastic rule that has been used to generate the soil properties layers is included as an illustration. It is based on the following assumptions:

1. There is no trend within a soil series.

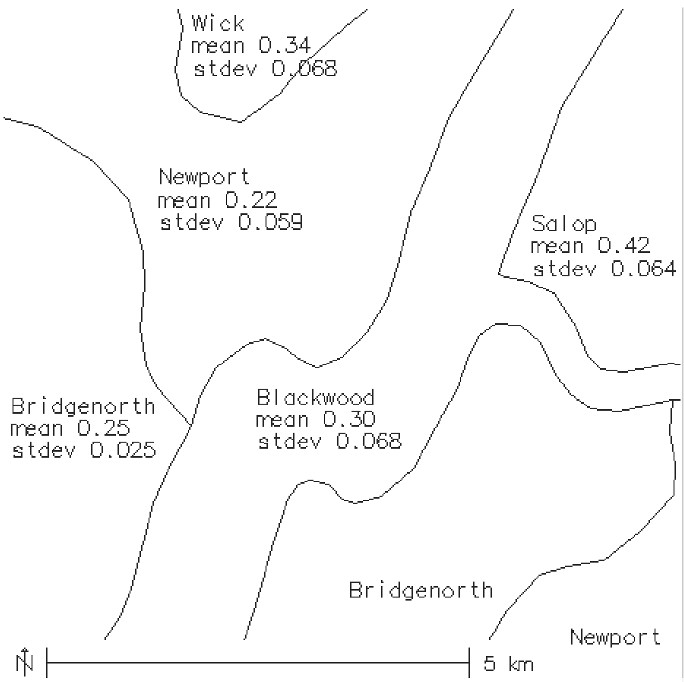
2. There is a weak spatial auto correlation over the grid size that has the form:
   p(d)=k^d (2)
3. Individual parameters are correlated with each other in a simple linear fashion.
4. Our measured points are augered samples not absolute values over the whole grid. We use these to estimate the spatial statistics and then generate representative values for the whole cell containing the pit.
5. Different realisations of the same stochastic set will give different values of yield loss and this will enable reliability figures to be applied to the loss estimates.

The standardised value for the most critical or least variable parameter can be calculated for each grid square:

v(i,j)=a**[**v(i-1,j) + v(i,j-1)**]** + bv(i-1,j-1) + c*n* (3)
v(i,j) is the standardised value for the property for cell row i column j.
a,b and c are parameters for each soil series and soil property to be established by method of moments using the available large scale data.
n is a random(0,1) number.

The standardised value is multiplied by the standard deviation and added to the mean of each soil series. Subsequent parameters are calculated using a similar method or generated from the single core distribution, whereby cross-correlation is accounte d for, but the very weak spatial auto-correlation is not specifically maintained

a



b

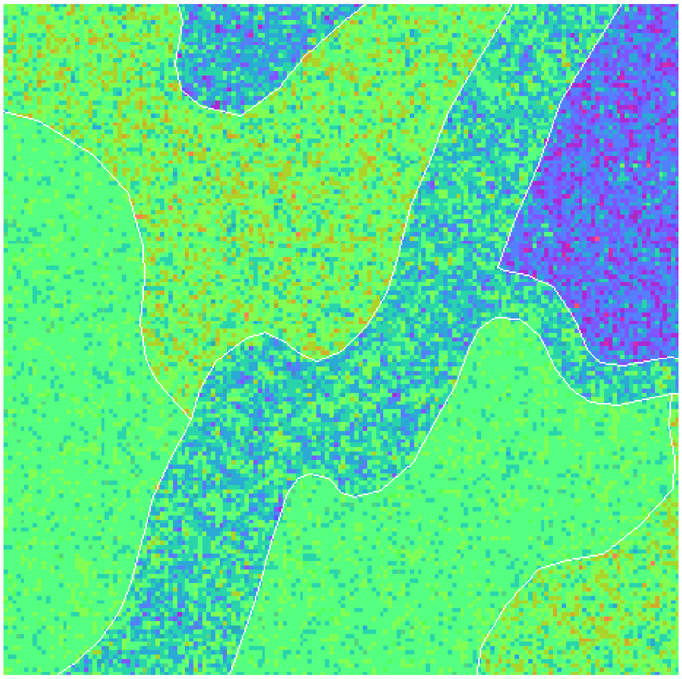


Figure 7,(a)an example of delineated soil series, (the real map cannot be shown because of copyright),(b)a realisation of the stochastic variation of saturated moisture content of the top horizon, this will be used as one of the raster layers in Figure 4.

This is of course a very simple formulation but it has the advantage that it is easy to implement using a map algebra facility, and serves well as a first step in handling soil series heterogeneity. To go any further may be unwarranted bearing in mind the accuracy of the model and other input parameters. More sophisticated approaches include those discussed by Gessier *et al*(1996, this publication)

The concept of a soil series is of course an attempt to handle soil heterogeneity but in a general purpose way. The question that remains to be answered is whether the use of soil series leads to sufficient specification of spatial variability of soil hydraulic properties to be of use for the current application. If the answer is no, then the use of some method, as described above, is called for.

Similarly, do the other sources of surrogate and secondary data describe the spatial variation adequately. The most obvious feature of the generated rasters is that they will have a spatially varying quality, depending on whether the soil series is tak en from the large or the small scale maps, which has to be taken into account when decisions are based on model outputs. The approach followed here is to go as far as is reasonable to capture spatial variability. Reasonable is defined in terms of the cos t in time and money of improving the quality of data with respect to the improvement in the overall modelling system and the estimates being made. Where accuracy is limited by other parts of the modelling system such as; the accuracy of other data within the model component, the accuracy of the model; the accuracy of the model components or the way the components are linked, then improving the parameters estimates for one particular subsystem are not justified.

CALIBRATION

An important stage in any numerical modelling is that of calibration (see Sorooshian and Gupta, 1995 for a good review). However the modelling system outlined above is not necessarily amenable to this process. Calibration can be applied to subsystems o f models, to the individual models or to the whole modelling system of GIS linked models.

Realistically, reaching an optimal set of inputs is probably impossible. However, obtaining a satisfactory set of parameters for a particular purpose and under a constrained set of conditions (in this case drying conditions with a falling watertable) m ay be possible, once it is accepted that the model will need to be recalibrated if used for some other purpose. The model system components can be calibrated individually, or the modelling system can be calibrated as a whole - again the soil system is des cribed by way of illustration

Calibrating the soil moisture component

Calibrating the soil moisture component is the most problematical aspect of the modelling system. Each point on the landscape is being modelled independently, i.e. lateral flow is not considered. However the input parameters are generated as spatially varying rasters. This a more realistic way to treat soil input parameters and, because the model is to be run at thousands of points across the landscape, specifying the input parameters individually is impractical.

The Shropshire Groundwater Scheme has a few sites at which the soil moisture regime has been monitored for a number of years. These form a good record against which to calibrate the model system. Here it is practical to vary the parameters individuall y. However once you have a set of parameters at a point the question is how to apply them over a large region. Once again we have the choice of treating the point data as representative of an homogeneous unit or attempting to take into account some of th e spatial variability of soils.

Hopmans and Guttierez-rave (1988), discussed the problem of calibrating the root uptake function of the soil moisture model SWATRE over a large area with a heterogeneous soil properties (NB, the root uptake function in SWMS_2D is based on the SWATRE fu nction). They found that using a monte-carlo analysis with a trial and error method to optimise parameters gave better transpiration results than with single site calibration. However, they also found a considerable error between model output and measure d transpiration in particularly dry years.

Fitting the model to monitored sites or point data that provide a very detailed description of the soil moisture profile through time but are spatially very sparse is one approach. An alternative is to use some measure with good spatial coverage but le ss detail in the profile. The obvious candidate for such a measure is remote sensing, particularly microwave images. The use of microwave remote sensing for soil moisture monitoring is a current research topic, see for example Rogows ki and Engman (1996, this publication). However this technique only interrogates the top layers which are primarily influenced by atmospheric processes. For this project it is the effect of the falling watertable that is important, and hence the use o f this approach should be treated with caution. Other possible candidates for an easily obtainable spatial indicator of moisture status include remotely sensed crop properties such as stress or temperature.

Calibrating the whole system

The object of the project is to estimate the environmental impact of groundwater drawdown on crop production. Thus crop yield is an appropriate measure against which the whole system can be calibrated. Though it is possible to gather spatially varying yield data, historically no data has been collected in this way within the Shropshire Groundwater Scheme. Some farm records are available, which give crop yields lumped either to the farm unit or in some cases to the field unit. Given the error and uncert ainty in the inputs, utilising this data for calibration is the most robust approach, as using a larger spatial unit acts to cancel out, but not remove, some of the smaller scale spatial error. There is also the potential for combining this data with remo tely sensed data to give a better estimate of the spatial variability of yield, through a simple overlay operation in the GIS.

Again the approach to calibration must be pragmatic and cost effective. Thus the method adopted is to calibrate both at the individual component level and at the modelling system level using all the available data either during the calibration or in th e verification, but not calibrating the system beyond what is required. Consequently a trial and error approach has been adopted.

SUMMARY

Linking large scale models to GIS with its data handling capabilities and its spatial manipulation functions makes it practical to apply the models over a large region.

For the project described in this paper experience has shown that the model system accuracy and hence its applicability to decision making are limited by the data availability and quality, together with the capability to calibrate the system rather tha n by the sophistication of the model components themselves. This is likely to be the case for any similar project unless a significant amount of time and money are invested in data collection.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Engineering and Physical Sciences Research Council and The Royal Academy of Engineering, and the co-operation of the National Rivers Authority Severn Trent Region.

REFERENCES

Binley, A.M., Beven, K.J., and Elgy J., (1989) A physically-based model of heterogeneous hillslopes. II. Effective hydraulic conductivities. *Water Resources Research* 22: pp 1531-1536.

Burrough, P.A., (1986) *Principles of Geographical Information Systems for land resources assessment.* Oxford Science Publications: Monographs on Soil and Resources Survey no 12.

Charnock, T.W., Hedges, P.D., and Elgy, J. (1996) Linking multiple process level models with GIS. *Application of Geographic Information Systems in Hydrology and Water Resources Management, Proceedings of HydroGIS'96*, Ed. K. Kovar and H.P. Nachtnebel, IAHS Press, Wallingford, UK. pp 29-36.

Evans, D.A., (1990) *Development of a conceptual model of the soil-moisture-plant sub-system of the hydrological cycle* PhD Thesis, The University of Aston in Birmingham.

Gessler, P., McKenzie, N., and Hutchinson, M., (1996) Progress in Soil-landscape Modelling and Spatial Prediction of Soil Attributes for Environmental Model s. *Proceedings of the 3rd International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modelling*, Sante Fe, New Mexico, January 21-25th.

Hedges, P.D., and Walley, W.J. (1983) A study of soil moisture losses due to the drawdown of a shallow water table. *Scientific Procedures Applied to Planning Design and Management of Water Resources Systems* Proceedings of the Hamburg Symposium A ugust 1983: IAHS publication no. 147: pp 489-499.

Hedges, P.D. (1989) *The impact on agriculture of the drawdown of shallow watertables*. PhD Thesis, The University of Aston in Birmingham.

Hopmans, J.W., and Guttierez-Rave, E. (1988) Calibration of a root water uptake model in spatially variable soils. *Journal of Hydrology*. 103: pp53-65.

Kiefer, E.M. (1993) A conceptual-stochastic model of unsaturated flow in heterogeneous soils. Journal of Hydrology. 143: pp3-18.

McDonald, M.G., and Harbaugh, A.L. (1988) A modular three-dimensional finite difference ground-water flow model. *Techniques of Water Resources Investigations of the United States Geological Survey*. Book 6, Chapter A1.

Rogowski, A.S., and Engman, E.T., (1996) Using a Decision Support System in a GIS Framework to Model Spatial Distribution of Soil Water. *Proceedings of the 3 rd International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modelling*, Sante Fe, New Mexico, January 21-25th.

Simunek, J., Vogel, T., and van Genuchten, M.Th. (1994) *The SWMS_2D code for simulating water flow and solute transport in two-dimensional variably saturated media, Version 1.2*. US Salinity Laboratory, Agricultural Research Service, U.S.D.A., Ri verside California: Research Report No. 132.

Soil Survey of England and Wales (1984) *Soils and their use in Midland and Western England*. Soil Survey of England and Wales: Bulletin no. 12.

Sorooshian, S. and Gupta, V.K. (1995) Model Calibration. *Computer models of watershed hydrology*: Singh, V.P. (ed), chapter 2, pp23-63.

van Genuchten, M.Th., (1976) A closed-form equation for predicting the hydraulic conductivity of saturated soils. *Soil Science Society of America Journal* 44: pp892-898.

Thomas W. Charnock
Research Student, Aston University
Department of Civil Engineering
Aston Triangle
Birmingham, B4 7ET, UK
Telephone: UK 0121 359 3611
FAX: UK 0121 333 3389
EMAIL: charnotw@sun.aston.ac.uk

John Elgy
Lecturer, Aston University
Department of Civil Engineering
Aston Triangle
Birmingham, B4 7ET, UK
Telephone: UK 0121 359 3611
FAX: UK 0121 333 3389
EMAIL: J.ELGY@aston.ac.uk

Peter D. Hedges
Lecturer, Aston University
Department of Civil Engineering
Aston Triangle
Birmingham, B4 7ET, UK
Telephone: UK 0121 359 3611
FAX: UK 0121 333 3389
EMAIL: P.D.HEDGESaston.ac.uk

Shawn Laffan

RAPID APPRAISAL OF GROUNDWATER DISCHARGE USING FUZZY LOGIC AND TOPOGRAPHY.

ABSTRACT

Models used for predicting groundwater discharge are invariably complex and data intensive, restricting their application to small catchments. To model over broad areas a simpler, minimalist approach has been adopted. The model, FLAG (Fuzzy Landscape Analysis GIS), uses a combination of only elevation data and a simple set of rules and is intended as a rapid appraisal tool for broad area predictive mapping of groundwater discharge sites. To use only elevation data the model assumes that all the factors affecting discharge are expressed in surface topography, and these may be accounted for by a set of geomorphometric filters. FLAG models within a framework of fuzzy logic, which reduces the uncertainty of the analysis by enabling continuous classification of the filters. Originally developed by Roberts et al (1996) FLAG has thus far produced results comparable to more complex models.

FLAG was applied to two agricultural catchments in New South Wales, Australia. The model performed well in one catchment but the depositional areas of the second catchment stretch the model assumptions to their limits. FLAG is currently too simple and requires some increase in complexity to better account for the variations found. It is, however, a promising approach to broad area predictive mapping of groundwater discharge.

A TOPOGRAPHICALLY BASED HYDROLOGICAL MODEL

In Australia many areas of groundwater discharge are prone to dryland salinity and are more susceptible to gully erosion. In turn this leads to waterway pollution and lost productivity. Consequently there is a need for predictive mapping of regions susceptible to waterlogging where evaporation leads to salinity. Groundwater discharge occurs at the small catchment scale and so models that operate at this level should be used. Models already exist that will do this but often these are too complex to be easily used at the regional scale.

THE MODEL

Most hydrological indices are derived from one of three sources: rainfall, geology and topography. Over large areas rainfall is important only temporally. Geology has a deterministic effect upon topography and in many cases will have some form of surface expression. Thus topography is one of the most important factors for large scale hydrological modelling, and has been shown to be a significant control of soil moisture in the landscape (Anderson and Burt, 1978; Burt and Butcher, 1985; O'Loughlin, 1981, 1986). If these factors can be quantified then simple hydrological modelling becomes possible. To use topography as the sole input it is assumed that topography is the sum of all factors that have acted on a point over time. This includes such factors as geology, land use, and climate. As a result many factors such as fractures and intrusive dykes may be identified through their influence on surface topography.

FLAG uses three filters calculated from a DEM that are used to create two discharge indexes of groundwater discharge, the Lowness Index and the Curvature Index.

1. The *lowness* filter assumes the ground water surface is a smoothed version of topography. Where the smoothed surface is above the real surface there is positive piezometric pressure and discharge will occur.

2. The curvature of topography in plan controls the convergence and divergence of soil moisture throughflow. Discharge is more likely to occur where sub-surface drainage paths bottleneck, concentrating ground water flows. This is the filter *concavity*.

3. Piezometric pressure is also regulated by contributing area. Within FLAG ground water contributing area is assumed to consist of the set of points that are continuously and monotonically uphill of a point. This may include points outside the catchment area as long as they are connected through an uphill path. Discharge will not occur where *lowness* or *concavity* are sufficient but contributing area is insufficient. This filter is named *upness*.

The Lowness Index is a combination of *lowness* and *upness*; and *concavity* and *upness* are used to create the Curvature Index. To create these indexes *upness*, *lowness* and *concavity* are converted to fuzzy sets (membership values (mu) are scaled in the interval [0,1]) and combined using the fuzzy set intersection operation where the result is the minimum of the two sets. This creates a fuzzy discharge index where values nearer to unity have a greater probability of being discharge. Explanation of fuzzy logic as it relates to FLAG is given in Roberts et al (in press).

Fuzzy logic carries the uncertainty through the modelling process in a reliable and simple manner. This is because it allows the modeller to treat data in a continuous manner. However, a fuzzy discharge map is difficult for end users to work with because it contains a continuous range of values. Consequently the final index is converted to a binary image of discharge and non-discharge using an alpha-cut operation. An alpha-cut creates a crisp set by promoting to full membership (mu = 0) all fuzzy values greater than a determined threshold (alpha), those below are demoted to non-membership (mu = 0). To do this the threshold value at which discharge occurs must be found.

In FLAG alpha is currently determined using a ground truthed sub-set within the area being modelled to use as a training set, effectively calibrating the model. Alpha is calculated using a comparison matrix similar to a statistical chi squared table (see Figure 1) to compare mapped discharge with predicted discharge.

Using this table Dowling *et al* (in press) calculated four diagnostic statistics (alpha-statistics) to calculate alpha, three are used here:

1. ***Accuracy*** . This is the total of correctly classified discharge and non-discharge cells divided by the total number of cells.

$$accuracy = \frac{a + d}{a + b + c + d} \quad (1)$$

2. ***Efficiency*** . This weights the total of correctly classified discharge cells in relation to total errors of omission (cells incorrectly classified as non-discharge).

$$efficiency = \frac{a}{a + c} \quad (2)$$

3. ***Power*** . This is the total correctly classified discharge weighted by errors of both omission and commission, and ignores correctly predicted non-discharge. Because values of *power* display little variation graphically, the square root of the quotient is used to improve the sensitivity of the display.

$$power = \sqrt{\frac{a}{a + b + c}} \quad (3)$$

Discharge occurs in only a small area of a catchment, thus correct prediction of non-discharge gives a high overall accuracy but does not indicate the success of the model. Consequently the optimal value of alpha is when *power* is at a maximum.

Another means of assessing the model is by comparing the total ground truthed discharge with the total predicted. This may be expressed as

$$proportion = \frac{a + c}{a + b} \quad (4)$$

In a catchment that has fulfilled its potential for discharge optimal prediction is when *proportion* is equal to 1. Although this remains a general assessment it will give a better indication when combined with *power*. For the majority of catchments discharge is not fully realised, and using a *proportion* of exactly one will not be predictive.

After the alpha-cut is performed the potential extent of discharge is predicted over the area modelled. Consequently the effectiveness of the model is ultimately dependant upon alpha.

MODEL APPLICATION

The Tout Park study area is located at 148�30'E, 34�30'S, approximately 15km north of Young, NSW. It consists of a north aligned central ridge with numerous first and second order catchments draining east and west. The study area is approximately 4.5km by 2.5km. Elevation ranges from 440m to 540m. Mean slope is 4.4�, and values range from zero to 12.3�. Geology is granitic on the west side of the study area and sedimentary on the east.

The contact is about 200m to the east of, and parallel with, the ridge. Numerous intrusive dykes are aligned with the ridge and there are few depositional areas within the study boundary.

The model was also run at the Mona Vale study site. Mona Vale is an easterly draining third order catchment situated at 147°30', 35°12', 15km south-east of Wagga Wagga. The catchment dimensions are approximately 2.5km by 2km, and total area is 228ha. Elevation varies from 210m to 340m. Mean slope is 6°, although slopes exceeding 25° occur in the headwaters of the catchment. Geology consists of steeply dipping interbedded Ordovician siltstones, shales and phyllites, with some schists and quartzites. The general strike is south-west, and the dip angle is 75° (Raymond, unpub.). Orthogonal jointing occurs along north-east and north-west axes (G. Tassel, pers comm, 21JUL94). Major land uses are cattle grazing and cropping, and tree cover is about 2%. Numerous contour banks are present and discharge and saline areas have been fenced off.

A cell size of 10m is used as it has been found to give gives an acceptable compromise for spatial accuracy with topographic indices (Zhang and Montgomery, 1994)

Concavity was calculated using Zevenbergen's and Thorne's (1987) algorithm. The curvature histograms for both DEMs are centred around an even curvature with long positive and negative tails. Because of this a sigmoidal membership function was used to create the fuzzy coverage.

A smoothed DEM was calculated using a 130m by 130m smoothing window. The use of this window size is to maintain some similarities with Roberts *et al* (in press) and does not necessarily reflect an optimal value. The original DEM was subtracted from this surrogate groundwater surface to create a grid of altitude differences. From this grid *lowness* was calculated using the linear membership function

$$\mu = \begin{cases} 0; & diff \leq 0 \\ \dfrac{diff}{\max(diff)}; & diff > 0 \end{cases} \qquad (5)$$

where *diff* is the resulting value after subtracting the two DEMs, and max(*diff*) is the largest difference in the grid.

Upness was calculated using a FORTRAN program written by D.W. Roberts for the original analyses of Dowling *et al* (in press) and used a linear membership function. The *upness* filter contains prominent lineations that could be due to the interpolation of the DEM or the algorithm used to calculate *upness*, however this should not greatly affect the model.

Calculation of the Curvature and Lowness Indexes was by the fuzzy set intersection operation. The resulting index values are thus the minimum of *upness* and either *lowness* or *concavity*, as this is the degree to which a point satisfies the conditions of discharge imposed by the model. Interpretation of discharge was by field inspection and from aerial photographs using changes in vegetation and obvious drainage features such as gullies and discoloured vegetation.

The results of both indices were compared with ground truth to calculate the alpha-statistics for discharge. *Upness* was then linearly rescaled between [0,1] by imposing a maximum threshold, after which *upness* was set to 1. This was done to exclude extremely large *upness* values from influencing the results. The indexes were then recalculated until *proportion* was near to 1 and maximum *power* was achieved. The membership value of the indices at this point was taken as the alpha-value for the discharge maps.

RESULTS

Tout Park	Curvature Index	Lowness Index
α	0.33	0.19
Accuracy	0.86	0.86
Efficiency	0.65	0.58
Power	0.66	0.64
Proportion	1.12	1.01

The Curvature and Lowness Indexes are shown in Figures 2 - 5. The upper areas of discharge have been excluded, although the model does coincide with a significant proportion of the mapped discharge. The similarity of the results is due to the dominance of *upness*.

Mona Vale	Curvature Index	Lowness Index
α	0.84	0.84
Accuracy	0.95	0.96
Efficiency	0.56	0.45
Power	0.53	0.54
Proportion	1.57	1.01

See Figures 6 -9.

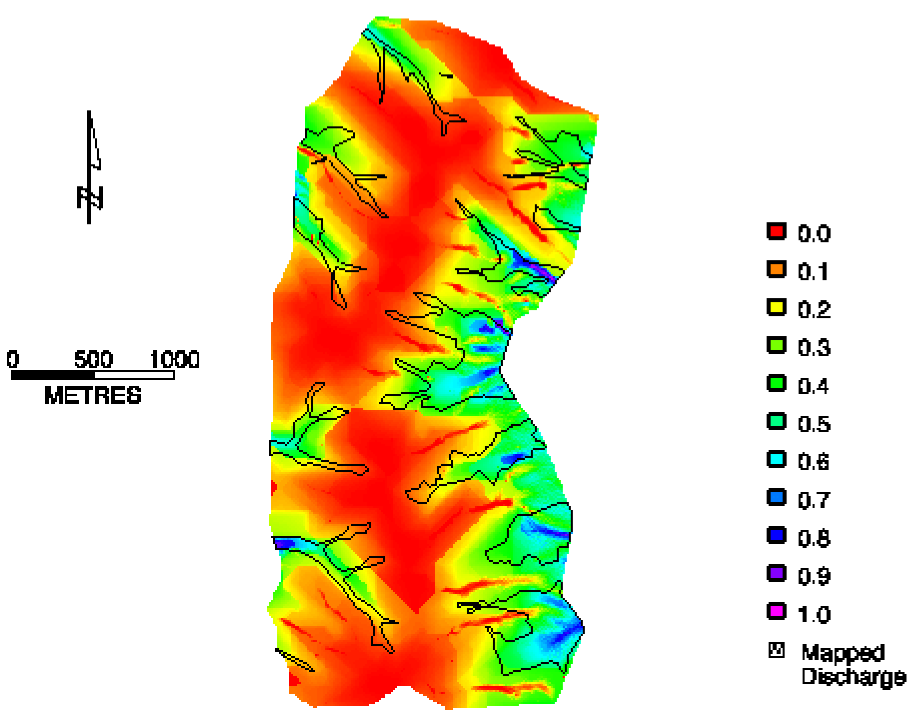


Figure 2: Tout Park Curvature Index

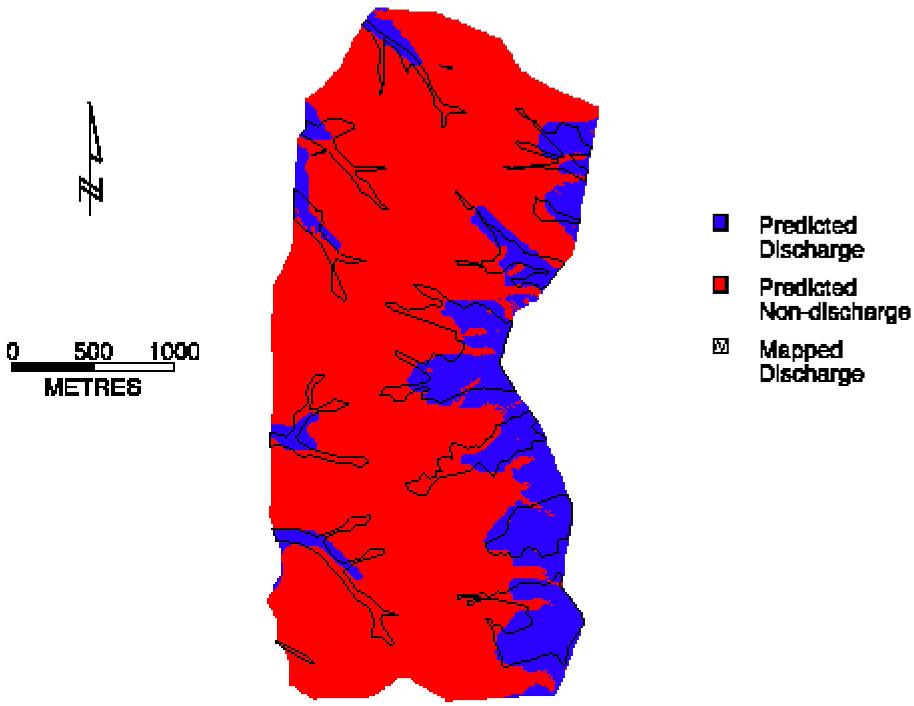


Figure 3: Tout Park Curvature Index alpha-cut alpha = 0.33

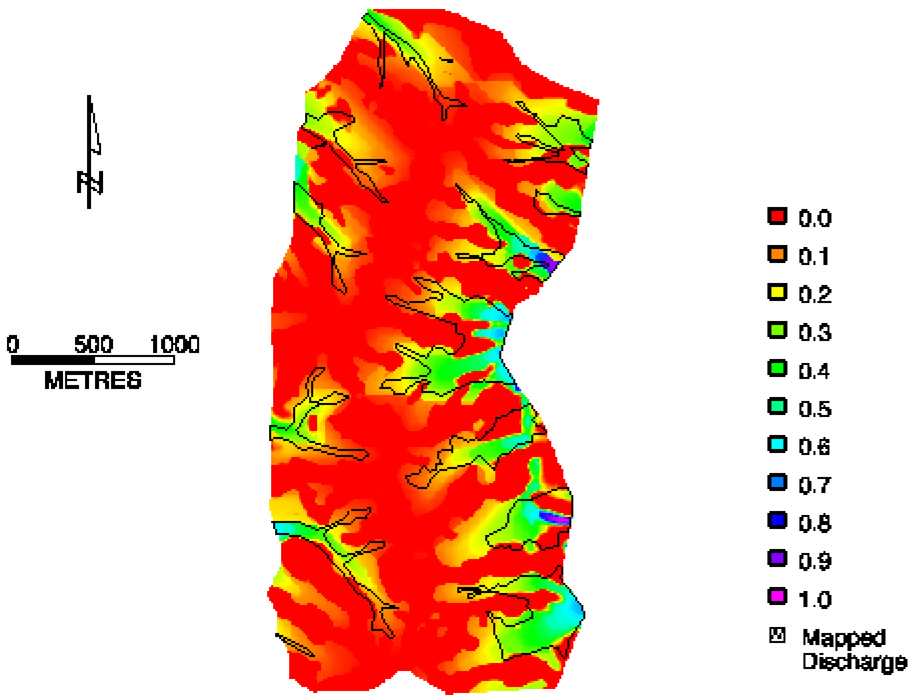


Figure 4: Tout Park Lowness Index

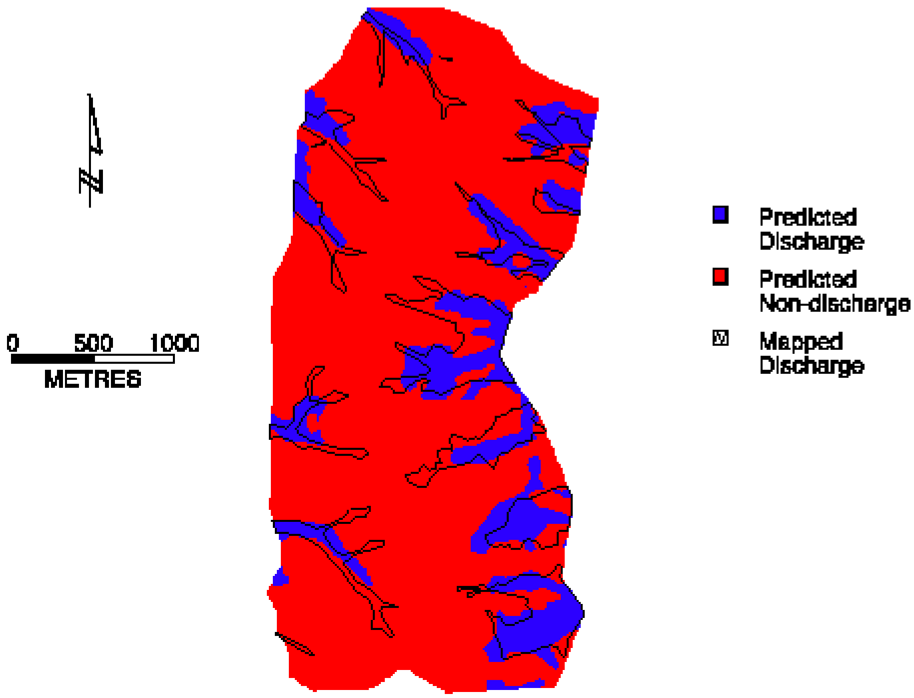


Figure 5: Tout Park Lowness Index alpha-cut alpha = 0.19

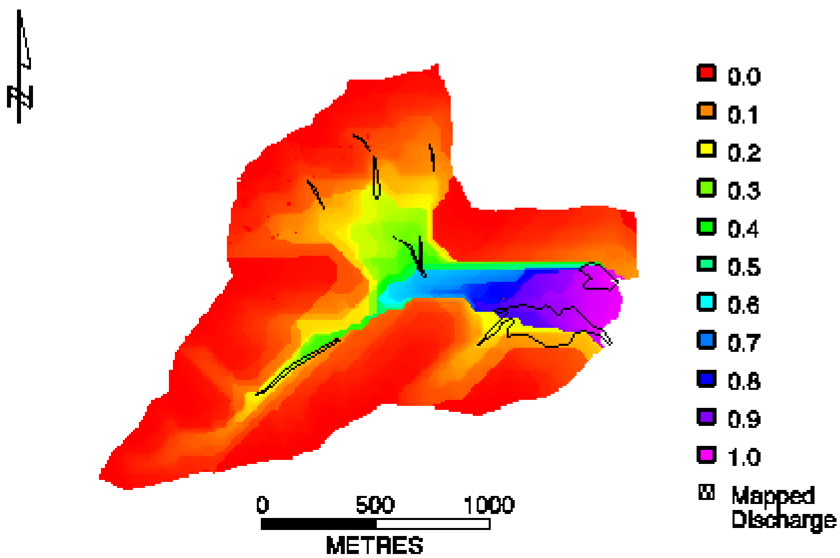


Figure 6: Mona Vale Curvature Index

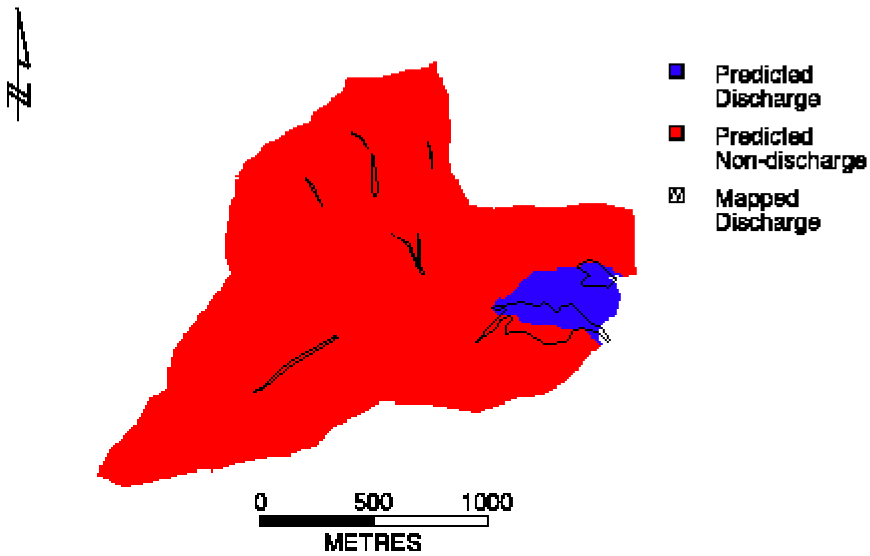


Figure 7: Mona Vale Curvature Index alpha-cut alpha = 0.84

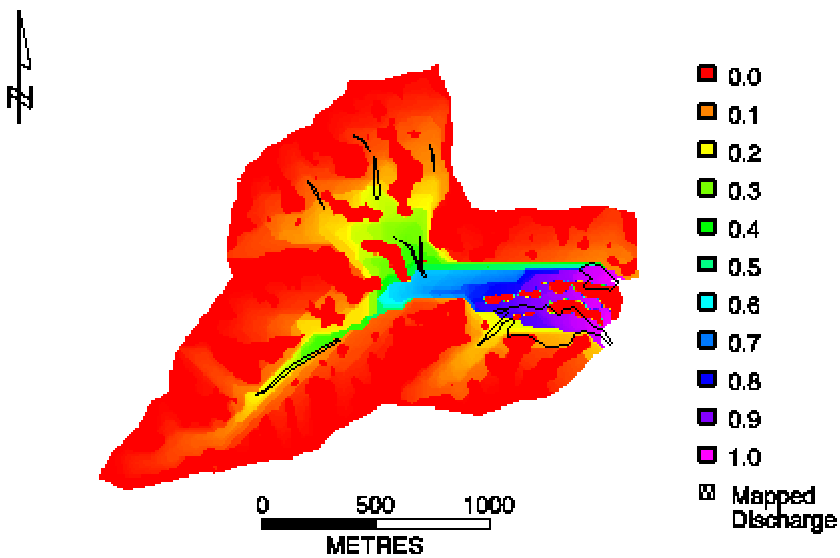


Figure 8: Mona Vale Lowness Index

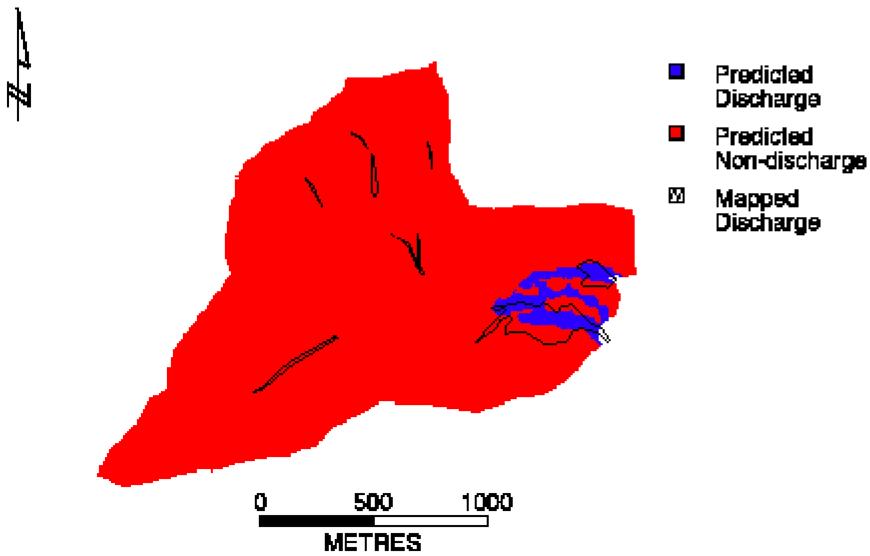


Figure 9: Mona Vale Lowness Index alpha-cut
alpha = 0.84

DISCUSSION

Tout Park appears to be satisfactory but the model assumptions have been greatly stretched at Mona Vale and the model is too simple to account for the variations present. To improve the model requires an increase in complexity. FLAG is intended as a rapid appraisal tool, and any improvement could be outweighed by the cost of its application.

The variations in the results are largely due to the different landforms of each catchment, which are due to geology, and this difference reflects the variety of discharge factors present in the landscape. Alpha is only valid for the landform in which it has been calculated. Also, the Tout Park study site contains less depositional areas than the Mona Vale site.

The results for Tout Park were better than at Mona Vale primarily because there is little depositional material within the study area. Also the local relief at Tout Park is greater than at Mona Vale and has trained a to more concave values.

At Tout Park the effect of the north oriented intrusive dykes has been differentiated by the model. The surface expression of these features is to constrict the gullies, creating bottlenecks that the *concavity* filter has differentiated and local depressions discriminated by the *lowness* filter. Tout Park is therefore a good example of where topographic models will work.

Assessing Model Accuracy

A comprehensive accuracy assessment of FLAG is extremely difficult, as the results are predictive. There is no true benchmark from which to assess accuracy. However, some indication of accuracy can be obtained by assessing the non-predictive mapping in the training set, or how much of the mapped ground truth was duplicated in the predictive image. The amount duplicated will give an indication of how well the model is at predicting discharge, assuming the alpha-cut is portable throughout the catchment. It is likely that some errors of commission will be associated with the non-predictive mapping, however these may be regarded as being either incorrect or predictive.

A simple assessment is easily obtained through the alpha-statistic *efficiency*. *Efficiency* weights errors of omission with correctly predicted discharge and, by ignoring errors of commission, is not affected by the predictive elements of the final map. The value of *efficiency* at the calculated alpha-value will give an indication of the model's accuracy as regards present discharge.

A Representative Training Set

Applying an alpha-cut throughout a catchment assumes that it is valid for that entire catchment. A fuzzy discharge index indicates the "potential" for discharge at a site. This forms the predictive aspect of the model and accounts in many ways for catchment differences induced by land management practices. To apply this potential across large catchments a representative alpha training set must be found.

Derivatives of surface topography may be used as indications of the type of topography, and make useful indices with which to define a training set. These derivatives are the general geomorphometric variables of Evans (1972): slope, aspect and curvatures, and are ideal for this purpose because they are easily calculated and are landform independent. A representative training set will be one that encompasses as much as possible of the distribution of plan curvature, profile curvature, aspect and slope present in the DEM. This will ensure that as many as possible of the full range of values will be used to determine alpha. Regardless, the training set must have discharge present in mappable quantities.

CONCLUSIONS

This study has been a test of the wider applicability of the model, and has highlighted some of the shortcomings of the model. However, when considering its simplicity and single data source, the model has performed satisfactorily. Other more complex models may be better at the small catchment scale, however many parameters are too costly to measure over broad areas and must be estimated or extrapolated from known data. It is questionable if their results will be any improvement on the minimalist approach used here. The low cost alone of the model makes it a viable option for large scale predictive mapping of groundwater discharge, as it is an order of magnitude less expensive than more complex models.

There are many other factors that will influence the results obtained using FLAG. Many of these may be surmounted by extra effort, but the purpose of the model is to provide a rapid and inexpensive appraisal of the landscape. Inclusion of geology is essential. Other topographic indices may be considered because they do not require an increase in data sources. These must be assessed before wider application.

Geomorphometric variables present a method of assessing catchments, but a method of determining which catchment to use as a training set remains labour intensive.

FLAG is thus a promising approach to modelling groundwater discharge.

BIBLIOGRAPHY

Anderson, M.G. & Burt, T.P. (1978) The role of topography in controlling throughflow generation, *Earth Surface Processes*, 3, 331-344.

Burt, T.P. & Butcher, D.P. (1985) Topographic controls of soil moisture distributions, *Journal of Soil Science*, 36, 469-486.

Evans, I.S. (1972) General geomorphometry, derivatives of altitude, and descriptive statistics, 17-90, in Chorley, R.J., *Spatial analysis in geomorphology*, Methuen, London, 393pp.

Evans, I.S. (1980) An integrated system of terrain analysis and slope mapping, *Zeitschrift fur Geomorphologie NF, Supplementband*, 36, 274-295.

Hutchinson, M.F. (1989) A new procedure for gridding elevation and stream line data with automatic removal of spurious pits, *Journal of Hydrology*, 106, 211-232.

O'Loughlin, E.M. (1981) Saturation Regions in catchments and their relations to soil and topographic properties, *Journal of Hydrology*, 53, 229-246.

O'Loughlin, E.M. (1986) Prediction of surface saturation zones in natural catchments by topographic analysis, *Water Resources Research*, 22(5), 794-804.

Rango, A. (1994) Application of remote sensing methods to hydrology and water resources, *Hydrological Sciences Journal*, 39 (4), 309-320.

Raymond, O.L. (Unpublished) Geology of the Wagga Wagga sheet, *Australia 1:100000 series, sheet 8327*, Australian Geological Survey Organisation (map only).

Roberts, D.W., Dowling, T.I., & Walker, J. (1996) FLAG: A fuzzy landscape analysis method for dryland salinity assessment, *CSIRO Division of Water Resources Tech Memo, in prep*.

Zevenbergen, L.W. & Thorne, C.R. (1987) Quantitative analysis of land surface topography, *Earth Surface Processes and Landforms*, 12, 47-56.

Zhang, W. & Montgomery, D.R. (1994) Digital elevation model grid size, landscape representation, and hydrologic simulations, *Water Resources Research*, 30(4), 1019-1028.

Shawn Laffan
Department of Geography,
School of Resource and Environmental Management,
Faculty of Science,
Australian National University,
Canberra, ACT,
Australia.
Phone: +61 6 2490348

Fax: +61 6 2493770
E-mail: shawn@dart.anu.edu.au

Toar T. Schell, Miguel F. Acevedo, Fred C. Bogs, James Newell, Kenneth L. Dickson, Institute of Applied Sciences, University of North Texas, Denton TX., and Foster L. Mayer Environmental Protection Agency, Gulf Breeze Research Laboratory, Gulf Breeze FL.

Assessing Pollutant Loading to Bayou Chico, Florida by Integrating an Urban Stormwater Runoff and Fate Model with GIS

ABSTRACT

This paper discusses the integration of Geographic Information Systems(GIS) and EPA's Storm Water Management Model (SWMM) as part of a watershed approach to assessing the ecological health of Bayou Chico, which is a sub-estuary of Pensacola Bay. Bayou Chico is the receiving water body of small, mostly urbanized watershed located in southern Escambia County, Florida. Both natural channel flow of precipitation and dry weather flow through a managed drainage system contribute to pollutant loading to the bayou. The transport and fate of pollutants are thus represented as water flow through these systems. A GIS database and the use of remotely sensed satellite images are combined to determine surface characteristics, storm drainage systems, area and slope of the watershed. These data layers are in turn linked to SWMM which mathematically represents these physical characteristics and uses this information to determine both runoff and pollutant loading. We discuss data required for SWMM, and its acquisition and use in the GIS. We will emphasize the GIS-model linkage with special attention given to model calibration techniques.

INTRODUCTION

In addition to the natural channelization of a given watershed, human endeavors have made the transport of water over and through a given watershed all the more impactive. Increased surface imperviousness concurrent with urbanization often result in the need to re-route stormwater flow. The introduction of stormwater drainage systems can expand the size of drainage area to a particular receiving water body. This results in the potential for increased pollutant loading to endpoints of such systems as pollutants are transported as a function of runoff. Physically based, spatially distributed models can be utilized to as a means of simulating dynamic runoff response for urban catchments. This is especially useful in cases where continuous measurements of discharge data are lacking , increasing calibration difficulties. Both structural and non-structural urban stormwater management strategies may be evaluated with greater confidence using physically based models, requiring a greater emphasis on database information (Meyer, 1993). The complexity of database information representing attributes for a given location necessitates the use of a GIS. Georeferenced features in a GIS are defined by points, lines or polygons, and may have several attributes attached to them each representing a different component of the hydrologic processes. The result of this is that model parameters can also be considered attributes for a particular map feature (Vieux and Needham, 1993). And that the integration of physically based mathematical models and GIS has become a common method for use in ordering the complexity of data necessary for the understanding of such systems (Moore et al., 1993).

Recently, efforts have been made by the University of North Texas (UNT) and the Environmental Protection Agency, to assess the ecological health of Bayou Chico: a small sub-estuary of Pensacola Bay, Florida. As part of a watershed approach to biological/chemical analysis of the bayou, a modeling component focuses on predicting the fate and transport of non-point-source pollutants in stormwater runoff entering the bayou. This is done through the integration of a physically based hydrologic model: the Environmental protection Agency (EPA) supported Storm Water Management Model (SWMM 4) (Huber and Dickinson, 1988), with a GIS: Arc/INFO 7.0.2 (ESRI, 1994). This paper describes criteria for model selection with a brief discussion of current SWMM/GIS interfaces available, as well as model calibration/verification techniques. We will also discuss a broad overview of SWMM integration with Arc/INFO based GIS analysis.

STUDY AREA

Located next to the city of Pensacola, Bayou Chico is the receiving water body of a small, mostly residential and urban watershed located in southern Escambia County, Florida. The watershed of Bayou Chico is defined primarily by the natural surface flow feeding its shoreline and contributing streams. Additional watershed area is included by the channelization of water through stormwater drainage elements that may not correspond to overland surface flow. We model watersheds Sub-bas 2 (109 ha.) for model calibration and SB-4 (556 ha.) for model verification. Both sub-watersheds comprise of primarily residential and commercial land use. Both drain into fresh water streams at the northeast and northwest portions of the bayou respectively. A large portion of the drainage area consists of a wetland/swamp which functions as a water purification system for the bayou (Pratt et al., 1993).

SWMM SELECTION

Model selection for the Bayou Chico study was based on the following criteria: (1) Must be physically based to take advantage of information derived from GIS analysis; (2) Model must easy to use, capable of simple to complex modeling with a minimum of program development; (3) Well documented with a history of use; (4) Model to be capable to model urban hydrologic processes the output of which can be exported estuarine fate and transport model; (5)must be cost effective.

First developed in 1971 under the supervision of the EPA, SWMM has a long track record that is well documented. Both the fortran code and executables are available at no cost, as is some documentation on file input formats for the various model modules or "blocks". Documentation while disparate, enables the user accessibility to the model's fundamental capabilities. A more detailed Users manual is available at a relatively small cost (see appendix A).

SWMM is a mathematical abstraction of the physical characteristics associated with an urban watershed. Topological characteristics including surface characteristics, stormwater drainage structures and gutters are described by the model. As such it is well suited to parameter input resulting from GIS analysis. The SWMM model is capable of a range of complexity in modeling from runoff in a single watershed with no pipe or channelization network, to that of a more complex system of watersheds and sub-watersheds, each feeding a pipe network with storage and treatment facilities. SWMM core programs are the modules or blocks: RUNOFF, TRANSPORT and EXTRAN. RUNOFF and TRANSPORT are capable or routing surface/groundwater flow and pollutant transport. Moreover, the transport block is capable of modeling dry weather flow and infiltration into sewer systems. The EXTRAN block while not capable of pollutant transport modeling is capable of complex hydraulic routing (Huber and Dickinson, 1988).

SWMM/GIS INTERFACE CONSIDERATIONS

SWMM has historically not had a user friendly interface. The integration of SWMM and Arc/INFO on a UNIX based platform has been recently been developed. SWMMDuet (Curtis, 1994) uses Arc Macro Language (AML) point and click programming as an interactive process for database management, GIS analysis and subsequent calls to run the SWMM program. The demonstration files which come with the program are designed to show how the database items for each geographic layer are to be structured. Pre-existing Arc/INFO coverages must be converted to this format with specific coverage and item names. The program when implemented should provide a seamless method of integrating the GIS with the model. Arc/INFO AML point and click widget programming tends to run slow on the Sun SPARC-2,5 and 20 workstations on which we tested the program . According to the author, the program runs under Arc/INFO 6.0 and a system with a IBM or DEC mainframe; however, there was some difficulty in getting SWMMDuet to run successfully under Arc/INFO 7.0 and a Sun Operating System. (Curtis, 1994). We were unable to successfully compile the Fortran77 code that came with SWMMDuet on our Sun SPARC-2 and 20 workstations and there was insufficient time on our part to debug the code to fit our system. For this reason, SWMMDuet was not used in our study, even though it would have been ideal.

The EPA has developed a windows based interface which allows for manual input of parameter values in a series of spreadsheet-like pages. Hydrographs and pollutographs for up to three outfalls along with measured data, as well as a plot of measured vs predicted data for a single outfall may be viewed after each run. New users and those wishing to use it as a screening tool will find the program invaluable. It also enables veteran users to systematically build the model with increasing complexity with model input displayed in a clear fashion. From within the Windows interface, there is no way to run in a batch mode, therefore changes in input must be carried out interactively. Users familiar with SWMM text format may find this method more cumbersome than SWMM run in batch mode both for sensitivity or calibration purposes. For screening purposes or for use as an initial data entry interface it is extremely useful. The DOS executable that is run from Windows is capable of batch-mode programming outside the Windows interface. An unofficial DOS based SWMM version is also available from the same source as SWMMDuet. It is recompiled to take advantage of extended memory and is constantly being improved (see appendix A).

MODEL/GIS INTEGRATION

Previous Studies

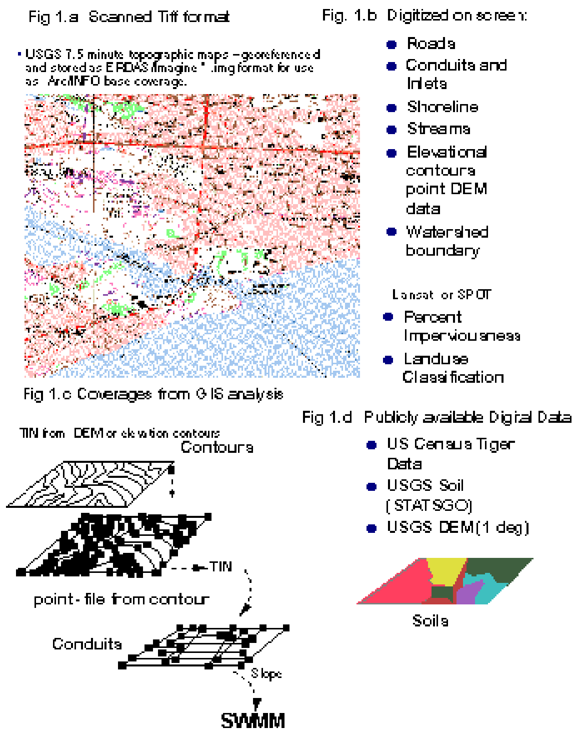
Our study builds on a previous study conducted by the Northwest Florida Water Management District (NWFWMD). This study used SWMM to calculate the mean flow volume and mean peak flow. These were then used to calculate pollutant load based on land use class. The use of GIS in this study was limited to land use analysis determining percent impervious surface (Pratt, 1989). Our approach seeks to use GIS to parameterize the model as much as possible, including watershed boundaries as well as stormwater routing and pollutant loading. The use of GIS analysis with the delineation of watersheds and the routing of storm runoff has been previously attempted by Stuebe and Johnston, (1991). Recent efforts have expanded the GIS/model integration to include water quality evaluation for urban watersheds. (Tim, 1992, Shea et al., 1993, Moore et al., 1993). Ross and Tara (1993) have identified five steps for model processing with GIS: (1) digital data gathering; (2) GIS operations; (3) model input data processing; (4) hydrological Simulation (5) output processing.

METHODS

Digital Data Gathering

Digital databases and their sources have been well documented (Moore et al., 1993). Such data range from freely available via anonymous file transfer (FTP) over the Internet, to manual digitizing data from hardcover maps (fig,.1.a,b,d).

Fig. 1 Digital Data and GIS Analysis - Input and Derived Layers for SWMM Parameter Integration.



The SWMM RUNOFF block can be broken into a hydrological component (Table 1.1), and a pollutant loading component. Both have parameters that may be retrieved from spatially based data. Pollutant loading is partially determined either by land use class or gutter length. Gutter length can be derived from digitized road layers or are available for some metropolitan areas in the U.S. Census Tiger Data files. Tiger roads are digitized at a scale of 1:100,000 and as such may have too many inaccuracies for more detailed modeling efforts. Maps from a local entity may serve this purpose better; however this data is available at little to no cost, and is already in a georeferenced digital format.

Table 1.1 SWMM RUNOFF hydrologic parameter data derived from GIS analysis.*

SWMM Parameter	Parameter Title	GIS Coverage Layer	Data Source	Format
Precipitation data Number and location of Rain Gages and associated rainfall	NRGAG REIN	Gageloci	Various Sources: NWS, on sight collection by local agency or private consultant.	Points with related Precipitation data
Channel/Pipe data Pipe name/number type (pipe, channel), width (m), length (m), inverted slope (ft/ft), left slope (ft/ft), right slope (ft/ft), mannings, depth when full (m), starting depth (m)	NAMEG NGP GWIDTH GS1 GS2, G6 DFULL GDEPTH	Conduits	Digitized from City of Pensacola Public Works maps On site survey determies attributes. Mannings derived from pipe/conduit material	Vector with relational data base
Inlets/Pipe network Inlet or channel for drainage pipe or channel routing	NGTO	Conduit-Flow	USGS 7.5 ' Topo. Maps, converted to TIN Arc/INFO TIN modeling routes flow from one element to another based on elevation to selected points (inlets)	Point data Vector data converted to TIN.
Watershed/Land use subcatchment data Subcatchment number, Area, Width, Slope,% Impervious area, % Pervious area, Impervious depression storage (mm), Pervious area depression storage (mm), Percent imperviousness w/ zero detention (immediate runoff)	NAMEW WAREA WW(3) WSLOPE WW(5) WW(6) WSTORE(7) WSTORE(8) PCTZER	Watershed, Basin Slope/Aspect PerviousLand Depress	7.5 min topographic map. Landsat TM, SPOT, ERDAS/Imagine Image analysis Software USGS 1 deg. Digital data, or digitized Vector polygon calculation within Arc/INFO	Raster or Vector contour maps polygons for land use and depression layers
Infiltration data (Green - Ampt) Average Capillary Suction (mm) Saturated Hydraulic Conductivity Initial Moisture deficit.	SUCT HYDCON SMDMAX	Soils	USGS STATSGO County Soil Survey Map	Vector -polygons with relational database

*Excluding Groundwater and Snowmelt

Image processing software such as ERDAS/IMAGINE 8.2 (1994) allow for the georeferencing of images interactively. RMS error is reduced by selecting ground control points (GCP) and moving them on screen (or dropping them) until an acceptable reduction of error is reached. This software was used to georeference Landsat TM (Fig 2.a) and SPOT images from August 1989 as well as perform a principle components analysis on the TM image using all of its spectral bands. The resulting image was then classified into land use categories and stored as an image file which was exported to Arc/INFO (Fig. 2.b).

Fig. 2-1 Landuse Analysis using Landsat Thematic Mapper (TM)

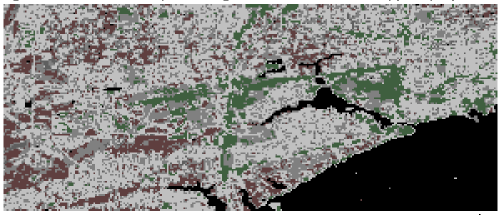
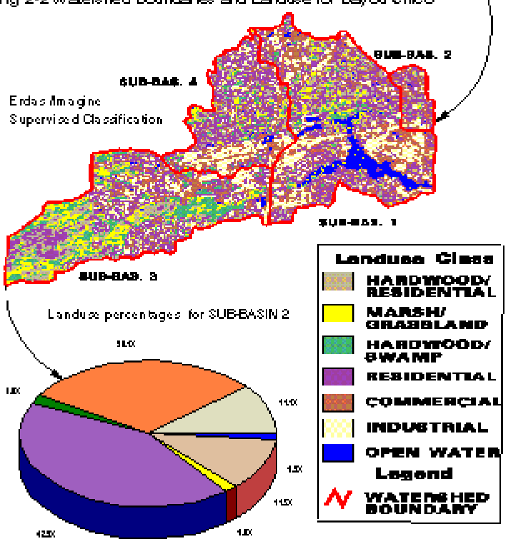
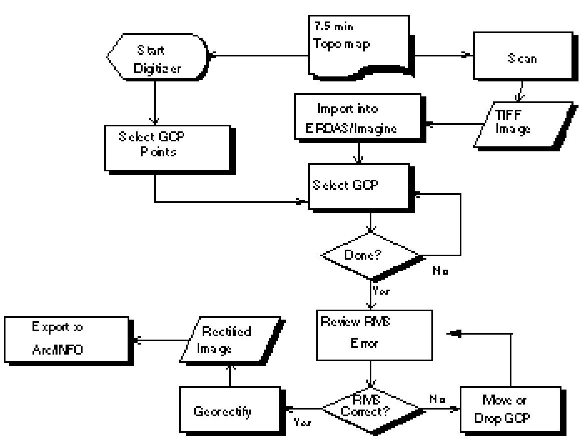


Fig 2-2 Watershed Boundaries and Landuse for Bayou Chico



We have used scanned 1:24000 scale USGS 7.5 topo maps in TIFF format and imported this to Erdas/Imagine and georeferenced this to the hardcopy version on a digitizing tablet (fig. 1.a). Using ARCEDIT module of Arc/INFO and the georeferenced topo image as the backdrop, data layers derived from the topo image were digitized on screen (fig. 3). Advantages of this method are: (1) Speed in processing - map is placed on digitizing tablet only once, reducing the possibility of map movement; (2) more control in reducing error; (3) zoom in/out function allows for more accurate digitizing, e.g. closely spaced elevation contour lines; (4) digitizing errors are more apparent; thus, are easily edited.

Fig. 3 Method for Georeferencing Raster Image Formats



GIS Operations

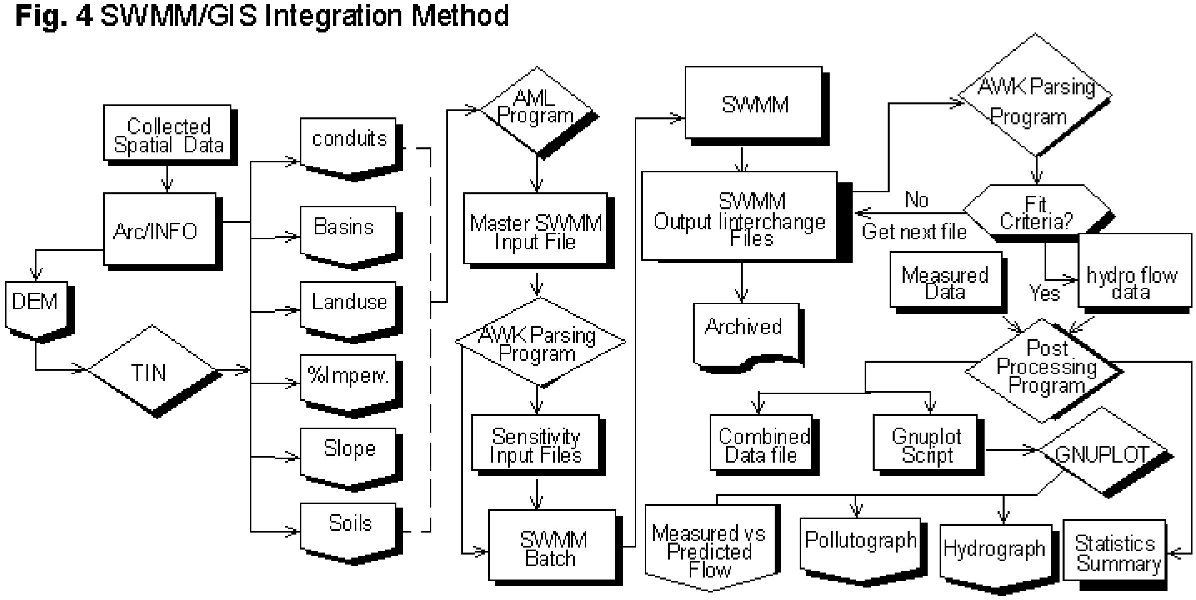
GIS operations consist of Data manipulation to derive parameter values from various georeferenced data layers. Coverages are either digitized or derived from the combination of several coverages. The Arc/INFO TIN modeling module is used to create a Triangulated Irregular Network (TIN) from DEM data or digitized points along elevation contours found on 7.5 minute Topographic Maps (Fig 1-C). When overlaid by channel networks or topology the TIN will serve to determine overland flow in the form a slope or aspect towards particular inlets or catchbasins as well as the slope of sub-watersheds and channel networks. Care should be taken when using TIN modeling in areas where the topology is relatively flat. Vieux and Needham (1993) have noted ambiguous results under flat conditions when parameters such as aspect and receiving cell location are being predicted. When encountering ambiguity, they opted for determining such parameters manually.

As a vector based GIS, Arc/INFO has tools that allow for querying of polygons for areal extent, perimeter, length, width and other associated attributes. AML programming is used to query named Arc/INFO coverages for SWMM parameter data and to write the information to an ASCII text in the SWMM input file format. GIS AML programming as planned is to be done on a layer by layer basis increasing with model complexity. To date only preliminary AML programs have been completed for a single watershed (sub-bas 2) with a single pipe network.

Model Data Input Processing

Model Input data consist of values derived from GIS analysis and precipitation data. Long-term simulations may require National Weather Service hourly or 15min rainfall data. A single storm event may also be modeled, and in such cases data from raingages may be entered manually or imported from an ASCII file. Output from the GIS analysis is in the form of a SWMM input template. Parameter values used for model calibration (e.g. impervious area Mannings friction coefficient) are written as a minimum value within the range for that surface type. This value then can be incremented in batch-mode and subsequently tested for contribution to model error. The Awk script parsing language program (appendix A) is used to create duplicates of this file each with a single parameter value changed by the incrementation value chosen for the particular parameter. The results of the Awk program are a user defined number of SWMM input files (runs) along with a batch file that calls these files to the SWMM executable (fig. 4).

Fig. 4 SWMM/GIS Integration Method



Hydrologic Simulation

The SWMM program when implemented in batch mode is designed to produce a series of runs with several parameters used to calibrate flow to a particular endpoint. A secondary Awk program is designed to take SWMM output files and parse out flow and pollution data. It also queries the output file for a lower limit for flow and boundary times for peak flow. We use Total runoff volume and time to peak flow as our benchmarks. Files which meet the criteria are passed along with measured data to a post-processing program for statistical comparisons with including regression and correlation analysis. Original files are archived in case they need further examination. This has proven to be an effective method for a quick sensitivity analysis for the calibrating parameters (fig. 4).

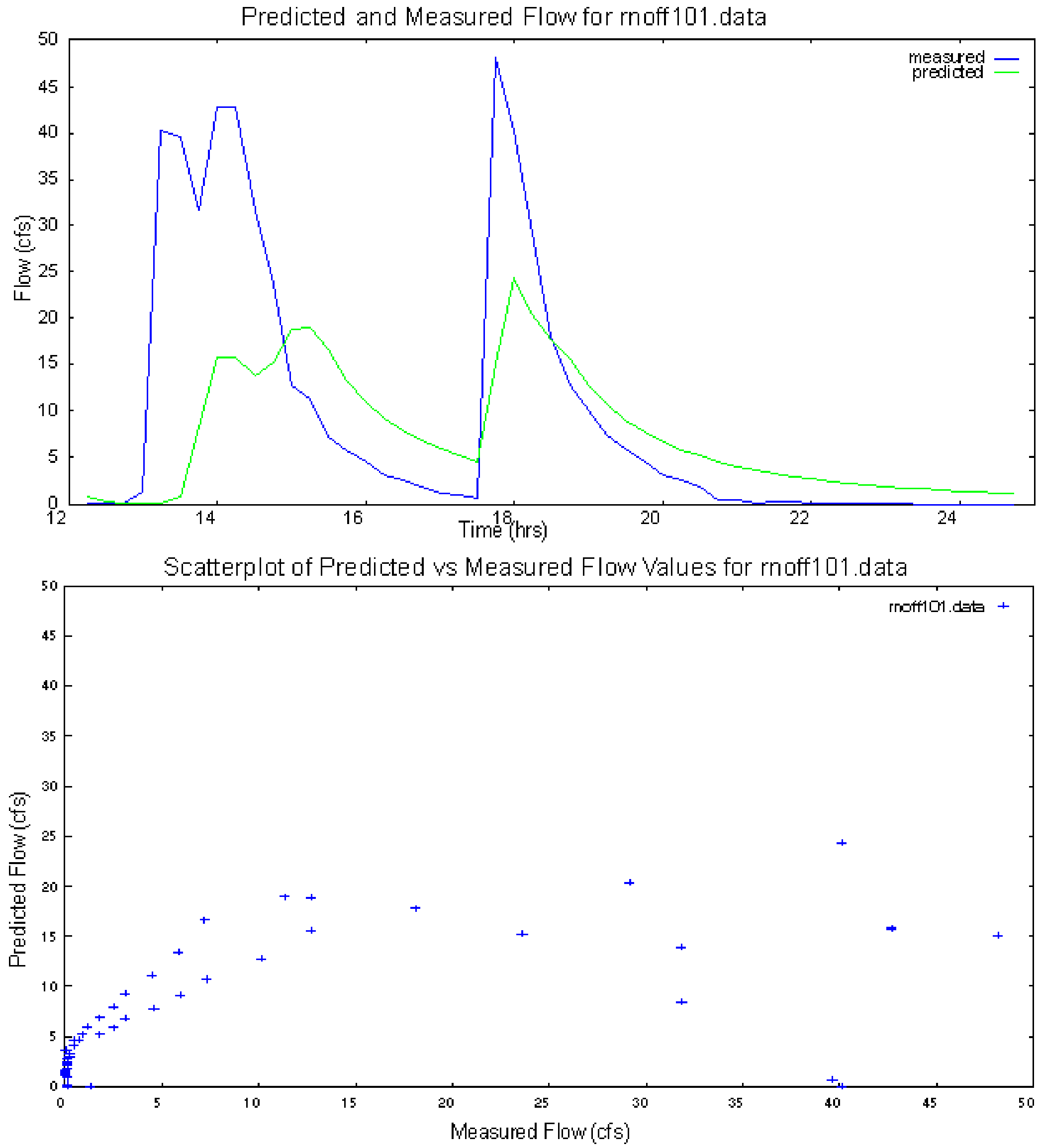
Preliminary runs have been performed on a simplified watershed with a single pipe network. Parameters from table 1-1 that have been used to calibrate the model have been the watershed width, pervious and impervious manning's coefficients, and three Green-Ampt infiltration parameters.

Output Processing.

The present graphic output from SWMM output files are designed for a line plotter and as such does not take advantage of modern printers. Traditionally one has had to import hydrograph and pollutograph data into a spreadsheet or other graphical program. Five goals were established for graphical outputs from SWMM: (1) View hydrographs, pollutographs and statistical graphics in an automated fashion; (2) View the results of multiple SWMM runs sequentially; (3) Produce graphs of presentation quality; (4) do this with minimal programing; (5) be cost effective

We take advantage of a publicly available graphics plotting program GNUPLOT, which has been developed for DOS, WINDOWS 3.x and UNIX (see appendix A for sources). We used a C program to read multiple combined data files, calculates regression and correlation statistics and also write a GNUPLOT script file to produce graphs for each run designated. Output files can be printed or viewed sequentially (fig. 5). This is an extremely effective method, both in terms of time and cost for viewing and publishing SWMM output.

Fig. 5 Graphical Output from Post-Processor Gnuplot Script



Conclusions

Preliminary results have shown that a SWMM GIS linkage for all spatially oriented attribute data is feasible. Results of the initial runs can be seen in fig. 5, notable are the time to peak and overall shape of the curve. Predicted total runoff volume (area under the curve) however, is much less than observed. Work still to be completed include the subdivision of the Sub-Watershed into drainage areas based on more than one inlet location. In this regard, results from initial TIN modeling have been ambiguous, primarily due to the flatness of the upper half of the Sub-basin. Pollutant loading has not been calibrated. Validation for the model will be performed on sub-basin 4 in the near future. To date completed tasks include the collection of all digital data, rudimentary AML programming to describe watershed parameters, and an automated screening and graphic interface with SWMM output.

APPENDIX A:

Internet Resources - no-cost programs

SWMM

The model is available in various formats. A windows 3.x version is available for remote download from the EPA World Wide Web site:

http://earth1.epa.gov/SWMM_WINDOWS/.

A newly compiled unofficial DOS version (4.31) of SWMM available from Oregon State University:

ftp.engr.orst.edu/pub/swmm/pc

This takes advantage of extended memory not available the official EPA version. (It doesrequire a math co-processor). This version is not supported by the EPA.

The unix based Arc/INFO-SWMM program

SWMMDuet

Also available at ftp.engr.orst.edu/pub/swmm/workstation along with a users Manual.

AWK

An awk version (awk, nawk, gawk) is found on most unix operating systems (including LINUX) A Dos version of Awk is available via anonymous ftp from:

http://www.acs.oakland.edu/oak/SimTel/msdos/awk.html

GNUPLOT

GNUPLOT is available from:

http://www.acs.oakland.edu/oak/SimTel/msdos/plot.html for both the dos andwindows versions.

A manuual is also available.

Dos: gpt35doc.zip (Image) 93/10/15, 600071 bytes

PostScript documentation for gnuplot 3.5 gpt35exe.zip (Image) 93/10/15, 552716 bytes

gnuplot 3.5: 2D/3D plots of data & fcns gpt35src.zip (Image) 93/10/15, 740888 bytes

MS Windows GUI version of gnuplot 3.5 Complete source of gnuplot 3.5 Windows: gpt35win.zip (Image) 93/10/15, 419933 bytes

---

References

Curtis, T.G. (1995). *SWMMDuet Users Guide.* Unpublished. Available from Oregon State University via anonymous ftp.engr.orst.edu/pub/swmm/workstation 42pp.

Huber, W. C. and Dickinson, Robert E., (1988), *Storm Water Management Model, Version 4: User's Manual.* Environmental Research Laboratory, Office of Research and Development, U.S. Environmental Protection, Athens, Georgia. 569 pp.

Meyer, S.P., et. al. (1993). "Geographic Information Systems in Urban Storm Water Management." *Journal of Water Resources Planning and Management*, 119(2), 206-228.

Moore, I.D. et. al. (1993) "GIS and Land-Surface-Subsurface Process Modeling." in Goodchild, M.F., Parks, B.O. and Steyaert, L.T. (eds.) *Environmental Modeling and GIS*, Oxford University Press, New York, pp. 196-230.

Pratt, T.R., et al. (1993). *Stormwater Assessment of the Bayou Chico Watershed, Escambia County Florida. Surface Water Improvement and Management Plan: A Comprehensive Plan for the Restoration and Preservation of the Pensacola Bay System.* Northwest Florida Water Management District (NWFWMD), Water Resources Special Report 93-7, Havana, Florida.

Ross, M.A., and Tara, P.D. (1993). "Integrated Hydrologic Modeling with Geographic Information Systems." *Journal of Water Resources Planning and Management*, 119(2), 129-140.

Stuebe, M., and Johnston, D. (1990). "Runoff Volume Estimation Using GIS Techniques." *Water Resources Bulletin*, 26(4), 111-116.

Tim, U.S., Mostaghimi, S. and Shanholz, V.O. (1992). "Identification of Critical Nonpoint Pollution Source Areas Using Geographic Information Systems and Water Quality Modeling." *Water Bulletin*, 28(5), 877-888.

Vieux, B.E., and Needham, S. (1993). "Nonpoint-Pollution Model Sensitivity to Grid-Cell Size." *Journal of Water Resources Planning and Management*, 119(2), 141-157.

Toar Schell,

Institute of Applied Sciences

University of North Texas

Denton Texas.

Email: schell@unt.edu

Phone (817) 565-4350

Theodore A.D. Slawecki, Ramesh K. Raghunathan, Victor J. Bierman, Jr., and Paul W. Rodgers

Modeling Resuspension of River Sediments using ARC/INFO

Please set your viewer to the width of the colored line for best viewing.

The Upper Hudson River, which is characterized by areas of sediment containing high concentrations of potentially harmful compounds, is currently the subject of an extensive study under the auspices of the Superfund Act. One aspect of this study is the assessment of potential risks to river flora and fauna from the remobilization of contaminated sediments through resuspension. Understanding and quantification of the risks will help involved parties to decide between remediation strategies such as no action, dredging, or in-place containment.

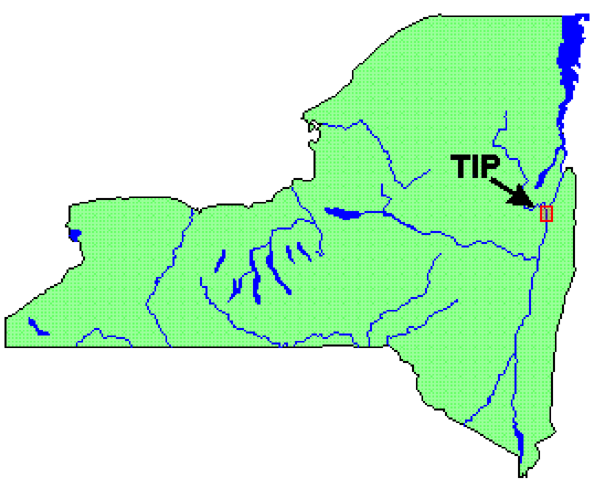
Part of the risk assessment is based on the modeling of bottom sediment scour during different flow conditions. Shear stresses at the sediment-water interface are predicted using a finite-element model and then imported into ARC/INFO as a point coverage. This point coverage is then combined with sediment properties and contamination coverages to predict depth of scour, resuspended sediment mass, and resuspended contaminant mass. Results are visualized with ARCPLOT or summarized as areawide totals.

The approach presented here:

- Provides a scientifically sound method for estimating sediment resuspension due to high-flow events.
- Accounts for spatial variability in model inputs such as sediment type and applied shear stress.
- Highlights resuspension "hot spots" for detailed consideration.
- Allows easy comparison of impacts of different flow scenarios.
- Is extendible to reflect better data and better scientific understanding of processes.

1.0 Introduction

The Thompson Island Pool, or TIP, extends for five miles along the upper reaches of the Hudson River in New York (Figure 1) from the northern tip of Rogers Island to the Thompson Island Dam. Organic compounds and metals entering the river above the TIP travel with the river flow, with portions attaching to solids suspended in the water which in turn may fall out of the water column and add to the sediments on the bottom. The materials attached to the sediments will be then be buried until a high-flow event resuspends the materials (Figure 2).



The remobilized organic compounds and metals in the water may have a significant effect on aquatic life in the Thompson Island Pool and downstream. Buried materials are believed to account for more than 98% of the contaminant mass in the TIP, so resuspension and remobilization during a high-flow event could represent a major source of pollutants to the water column.

Long-term monitoring data for flow and suspended solids suggest that sediment scour occurs at a flow threshold of about 11,000 cfs. Since this threshold is easily exceeded during moderate- to high-flow events in the TIP (the mean daily flow at Rogers Island is approximately 5000 cfs), a detailed GIS-based study has been undertaken to estimate the risk of sediment scour and associated contaminant remobilization during large flow events.

2.0 Approach

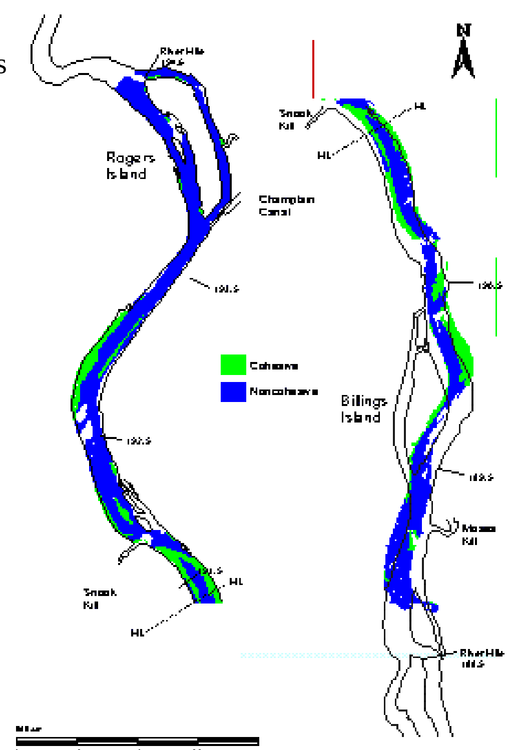
Estimating the total mass of contaminant remobilized during flow events requires three sets of calculations, as briefly described below.

2.1 Hydrodynamics calculations

The RMA-2V finite element model, developed by the U.S. Army Corps of Engineers, was used to model flow fields in the study area under conditions corresponding to low flow, springtime flow, and exceptional flow events such as 5-year and 100-year high flow events. Bathymetry data provided in ARC/INFO coverages were combined with elevation data digitized from USGS quad sheets to provide an accurate depiction of the river channel and flood plain areas for use in the model. A commercial software package, FastTabs by the Boss Corp., was used to help develop and evaluate a 2-D 6,000 node finite element grid (Figure 3) on a Pentium-equipped microcomputer running MS-DOS 6.22 and Windows for Workgroups 3.11. Model-predicted shear velocity at steady-state conditions was predicted throughout the study area.

2.2 Sediment resuspension

The bed sediments in the TIP range from coarse non-cohesive areas to fine-grained sediments in depositional areas along bends and shore lines (Figure 4). The area of non-cohesive sediments is approximately five times larger than the area of cohesive sediments.



In this study, resuspension was calculated differently for cohesive sediments (fine grain, high clay content with extensive interparticle effects) and non-cohesive sediments (no interparticle effects). Resuspension of cohesive sediments was modeled directly using an erosion equation proposed by Lick [Lick 1994] which provides a depth of scour, while resuspension of non-cohesive sediments was estimated using the Ackers-White formulation [Ackers 1973].

The Ackers-White formulation predicts the cross-sectional transport of non-cohesive solids transported in the water column, which in this application is related only empirically to resuspended mass. Mass transport in the water column is integrated at 24 equally-spaced lateral transects, and the median transect result is used to represent water column concentrations of non-cohesive sediments throughout the

TIP.

2.3 Contaminant resuspension

The amount of attached contaminant resuspended with the sediments is calculated by multiplying the total mass of sediments resuspended during the modeled event by the concentration in the sediment of the material of interest.

3.0 Use of ARC/INFO

ARC/INFO version 6 running on a Sun SPARCstation 20 was used for data management, resuspension calculations, and for visualization of model results (Figure 5).

3.1 Data management

One important role for ARC/INFO in this project was the management of spatial data for the application. Available data sets included:

- ARC/INFO polygon coverages for sediment types based on side-scan sonar investigations.
- ARC/INFO point coverages for contaminant sediment core sampling locations.
- Tabulated analytical data for sediment cores.
- Tabulated bathymetry data (transects every 100 feet, data spacing about 5').
- AUTOCAD drawings depicting shorelines and islands.

The analytical data was converted into an INFO file linked with the sediment core point coverage, and the bathymetry data was filtered and entered as a point coverage. In addition, Limno-Tech digitized elevation information from USGS quad sheets for the floodplains adjacent to the TIP.

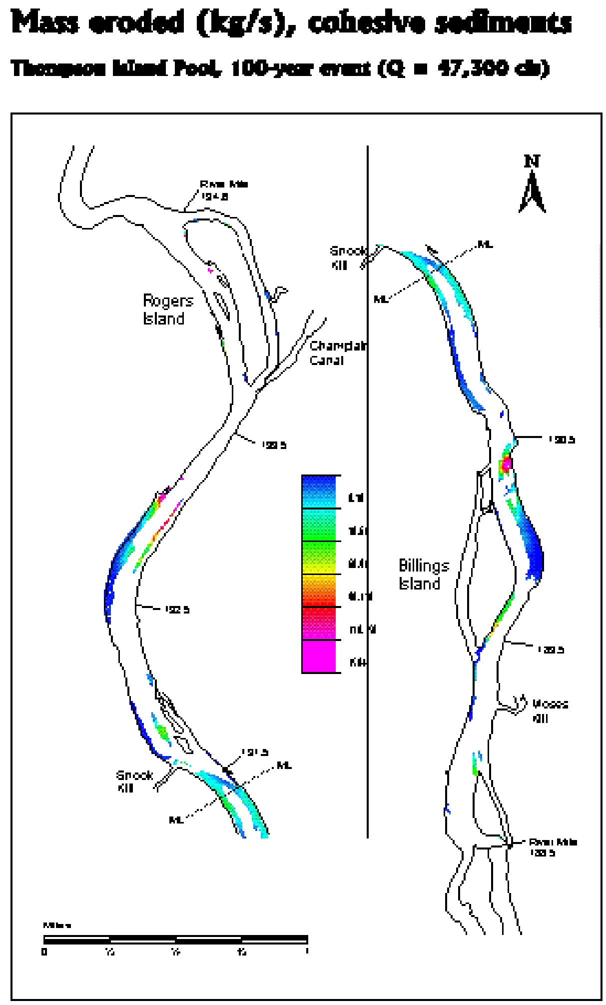
3.2 Resuspension calculations

This section describes the procedure for estimating the resuspended mass of cohesive sediments and attached contaminants (non-cohesive sediments are not discussed because the modeling approach is still under discussion). Example AML for this calculation is included in the Appendix.

1. Export bathymetry and floodplain elevation data from ARC/INFO to RMA-2V and develop finite element grid. This step occurs only once.
2. Determine shear velocities at steady-state conditions in the TIP for the flow condition of interest, and import the results into ARC/INFO as a point coverage.
3. Create a 10m grid of shear velocities using the imported point coverage.
4. Apply Lick's erosion equation to create grids characterizing cohesive sediment mass resuspended and depth of scour (Figure 6). Based on statistical analysis of laboratory and field data, Lick proposed an erosion equation of the following form which approximated his experimental data:

   epsilon = a0 / td^n * ((tau - taucrit)/taucrit)^m

   where epsilon is the total amount of material resuspended (g/cm^2), td is the time after



deposition, tau is the shear stress, taucrit is the critical shear stress, and and a0, n, and m are empirical constants. The depth of scour can be calculated as:

zscour = epsilon / Cbulk

This equation has been applied and results validated to several rivers (e.g. Fox River, WI; Detroit River, MI; Buffalo River, NY).

5. Use kriging to smooth the resuspension grid, then clip to TIP boundaries and mask out islands.
6. Multiply resuspended sediment mass by contaminant concentration to create a grid of contaminant mass remobilized (Figure 7).
7. Convert result grids to point coverages and operate on the PAT to get systemwide totals.

Preliminary results show that, as expected, the mass of solids and contaminants eroded increases as the magnitude of the flood increases. We found, however, that many "hot spots" - areas of sediment with high contaminant concentrations - are not likely to experience significant erosion due to their location in areas of the river bed subjected to relatively low shear velocities. This finding was aided by the fine-scale GIS approach used in this study.

3.3 Visualization

ARC/INFO was an excellent tool for viewing model results, giving the capability to produce graphics which accurately conveyed large amounts of information with some style. Graphic depictions of other spatial information was also useful in gaining an understanding of unique site features, such as the spatial distribution of sediment types and contaminants throughout the TIP, and for diagnostic examination of intermediate calculations like shear stress. The only significant difficulty faced in producing graphics was the development of an effective standard layout which could fit the entire area of interest on tabloid (11"x17") sheets for inclusion in bound reports.

4.0 Future Development

Several areas have been identified for the future enhancement of the modeling system described here, including the development of more accurate formulations for non-cohesive sediments, inclusion of vertical variability in contaminant and sediment properties, and generalization of the AML for use on other sites.

4.1 Non-cohesive sediments

One of the major concerns in the calculation of resuspension is the treatment of non-cohesive sediments. The depth of scour is not directly calculated; it is instead estimated from a sediment transport term based on the Ackers-White formulation. An alternative, more conceptual, statistically-based method is under consideration. The expected depth of scour would be directly calculated based on the particle size distribution in the bed, the critical shear stress by particle type, and the observed shear stress. This approach will lead to a more credible process parameterization by minimizing the number of calibrated parameters.

4.2 Vertical variability

The current implementation of the sediment resuspension calculation acts does not use all of the available information about sediment stratification and vertical distribution of contaminants in the sediment. This may lead to inaccuracies when the calculated depth of scour exceeds the thickness of the first layer of sediments, exposing a new layer with different characteristics which may not resuspend at the same rate. Similarly, contaminant concentrations in the sediments are represented in the model by the depth-averaged concentration, even though they actually change with depth. In addition, new data has been collected providing a more accurate description of depth-varying characteristics. We plan to develop additional AMLs to model the resuspension of sediments and remobilization of contaminants by layer, thereby increasing the accuracy of the model and taking advantage of newly available data.

4.3 Generalization

Most of the AML and linkage code developed for this application were created on a somewhat ad-hoc basis to meet project-specific needs. It is therefore not highly suitable for re-use on other sites. We plan to restructure the existing code to make it easy to re-use and to improve the user interface through the judicious application of menus and forms.

5.0 Conclusion

The application of state-of-the-art formulations for sediment resuspension offers an effective tool for predicting contaminant resuspension under different flow conditions. Combining these formulations with GIS allows their application in heterogeneous systems where contaminant concentrations and sediment type and grain size vary spatially in two or three dimensions. GIS also helps the modeler by providing effective depictions of site characteristics and model results.

Finally, the results of the application described here can be used to model contaminant resuspension effects on aquatic life forms and to decide what remediation strategies are necessary and appropriate for consideration.

Appendix - Example AML

The following is excerpted from the AML code used to calculate the mass of resuspended cohesive sediments and the associated contaminants.

```
/* Create 10 m grids directly from shear velocity
pointgrid t100 t100-sv shearvel;10;y;nodata

/* Calculate shear stress, depth of scour, mass resuspended
grid
t100-str = sqr(t100-sv) * 1000 * 0.929
t100-cm = 3.048 * 3.048 * (1.3 / 1000 / 7) * pow(t100-str - 1,2.5) * 100 * 100 / 1000
t100-csc = t100-cm / 3.048 / 3.048 / 715.5 * 100
quit

/* Convert results back to points
gridpoint t100-cm t100-cmp mass
gridpoint t100-csc t100-cscp scour
gridpoint t100-str t100-strp stress
kill t100-cm
kill t100-csc
kill t100-str

/* Use kriging to smooth resuspension estimates, then clip to TIP
/* and mask out islands.
grid
setcell 10
t100-amk = kriging(t100-cmp,mass)
t100-amk2 = selectpolygon(t100-amk,tip.seg)
t100-amk3 = t100-amk2 + isl-grid
t100-cmk = t100-amk3
kill t100-cmp
kill t100-amk
kill t100-amk2
kill t100-amk3
quit

/* Get contaminant resuspension by multiplying mass by concentration
grid
setcell 10
t100-cpcb = t100-csck / 30.48 * pcbsurfgr * 3.048 * 3.048
quit
```

Acknowledgments

The work described here was supported by U.S. EPA Contract #68-S9-2001 for Hazardous Waste Remedial Services under the U.S. EPA Region II Alternative Remedial Contracting Strategy. The authors acknowledge the assistance and cooperation of TAMS Consultants, Inc., Gradient Corporation, and U.S. EPA Region II.

References

Ackers, P. and W.R. White. 1973. *Sediment transport: new approach and analysis.* J. Hydr.Div., ASCE, 99(HY11), pp. 2041-2060.

Lick, W., J. McNeil, Y. Xu, and C. Taylor. 1994. *Measurements of resuspension and deposition in rivers.* Draft report to U.S. EPA Large Lakes Research Station.

Limno-Tech, Inc. 1995. *Draft Phase 2 Preliminary Model Calibration Report for the Hudson River PCB Reassessment RI/FS.* Report prepared for TAMS Consultants, Inc., and Gradient Corporation under EPA Contract #68-S9-2001 and submitted to U.S. EPA Region II, New York, NY.

Theodore A.D. Slawecki, *Computer Manager* (tad@limno.com)
Ramesh K. Raghunathan, Ph.D., *Project Engineer* (rkr@limno.com)
Victor J. Bierman, Jr. , Ph.D., *Senior Scientist* (vjbierman@limno.com)
Paul W. Rodgers, Ph.D., *Vice President* (rodgers@limno.com)

Limno-Tech, Inc.
501 Avis Drive
Ann Arbor, MI 48108

World Wide Web: http://www.limno.com
E-mail: Postmaster
Telephone: (313) 332-1200
Facsimile: (313) 332-1212

Thomas Goddard, Len Kryzanowski, Karen Cannon, Cesar Izaurralde, Tim Martin

Potential for Integrated GIS-Agriculture Models for Precision Farming Systems.

ABSTRACT

Precision farming aims to optimize the use of soil resources and external inputs(fertilizers and herbicides) on a site specific basis. Precision farming takes advantage of rapidly evolving GPS technology together with electronic sensors and controllers to monitor crop response under variable inputs and landscape position. Objectives of this study were: (i) to discover the soil-landscape-input relationships governing crop yields in characteristic Alberta landscapes, (ii) to test the performance of an agricultural simulation model using site-specific data, and (iii) to develop a method to analyze high-resolution data by linking the model to a GIS. Crop yields were monitored during 1994 and 1995 at four sites in Alberta using a high-precision 3-D DGPS. The Erosion-Productivity Impact Calculator (EPIC) model was run on a sub-field, site-specific basis using soil pit information and two levels of climatic data. The crop growth routines were compared against two years of yield maps obtained from 40 - 100 ha fields. A program was designed and built to couple the EPIC model with the GRASS-GIS. Our results show the potential to depict and analyze the variance in crop yield, leaching potential, erosion risk and economics on a farm field scale. The application of this approach in precision farming would allow for optimizing the use of resources on a site-specific basis thereby contributing to minimize detrimental environmental impacts such as nitrate leaching or erosion.

INTRODUCTION

Farming systems are continuing to change in response to economic, technological and social trends. Farming practises are becoming questioned not only by farmers but the public at large. Concerns are profitability and environmental impact. Farmers look to adopt new technology. In Canada, the trend over the past two decades has been to reduced tillage systems (direct seeding) and increase use of fertilizers and herbicides. The cost savings of reduced tillage does not outweigh the added cost of crop chemicals so profit margins have remained narrow. Our labour costs are high and our natural resources finite so we must utilize technology to maintain competitiveness in a world market. Developments in sensor- controller technology, computers and positioning systems now bring new opportunities for farm management (Goddard et al. 1995).

The development of the publicly available global positioning system (GPS) has opened new doors in opportunities for spatial data. This is a passive positioning system from a constellation of 24 orbiting radio-navigation satellites. They provide continuous position data provided the receivers have a line of site access (or nearly so) to the satellites. Positioning can be two or three dimensional in real-world coordinates.

Differential GPS (DGPS) uses a stationary monitor receiver to calculate the difference between the true position and the determination of a position from the satellites for a point in time. This allows the differential correction of a roving (remote) receiver so that the errors, induced as part of the military declassification, can be removed. This allows the positions of a roving receiver to have a three dimensional accuracy of 20 cm or better (Lachapelle et al. 1994). The use of a radio modem allows the transfer of differential corrections and corrected position determinations in real-time. The combination of accuracy and real- time determinations presents possibilites for guidance of farm equipment and the development of accurate digital elevation models (DEMs) as a by-product of other GPS aided farm operations (e.g. yield mapping). Experience in Alberta has shown that harvesting operations on a 80 hectare field can also yield 100,000 elevation measurements. Subsequent computer processing and terrain analyses can provide useful information to augment yield map interpretation.

Detailed yield map interpretation combined with terrain analysis from high quality DEMs and site specific soil sampling will provide new opportunities for the use of integrated crop models. Modeling landscapes and crops may negate the desire of expensive grid sampling which is the current recommendation to those entering the practise of precision farming.

One model that holds some potential for this application is the Erosion Productivity Impact Calculator (EPIC). Using either simulated weather (based upon monthly parameters) or daily records it will predict crop growth under a variety of conditions. It contains a soil model, tillage models, water erosion models (variations of USLE), a wind erosion model, hydrologic processes and pesticide movement, all dynamic on a daily time step. Economic parameters are included in the model as static parameters. EPIC has been shown to be effective for a variety of purposes and recent interest has been expressed on the Canadian prairies to evaluate EPIC with long-term research plots (Moulin and Beckie 1993; Toure et al 1995).

A research project was initiated in 1993 in Alberta to look at the application of GPS and related technologies at a farm scale. The objective was to use DGPS to allow yield mapping of fields, site specific sampling and variable rate fertilizer application. The goal was to develop the ability to apply the optimal rate of fertilizer for each part of a field for maximum economic yield and minimum environmental impact. The project presented the opportunity to examine the use of integrated crop models for application in site specific management systems.

A second project utilizing the same sites and equipment was initiated in 1994. The objective of this project was to examine an alternative approach to grid sampling of landscapes in order to reduce costs to farmers. The project would take advantage of detailed terrain data provided by high precision DGPS and integrated crop models to study the development of optimum agronomic management on a site specific basis. The study would be further aided by the development of software to integrate the EPIC model with the GIS, GRASS in order to provide a prototype to provide agronomic recommendations and risk assessment on a site specific basis across a field.

METHODS

Four farmer cooperators were selected in a north-south line through Alberta in order to encompass a range of soil and climate conditions (approximately 49-54 degrees N and 112 degrees E). One test field was selected on each farm. One site was irrigated (57 hectares) and the other three were dryland fields ranging from 32 to 81 hectares in size. Field characteristics were typical for the area (Table 1).

Table 1. Field characterization for three dryland sites in Alberta, Canada.

	HUSSAR	STETTLER	MUNDARE
Frost Free Days	115	115	100
Annual Total Degree Days (above 5C)	1550	1450	1400
Mean Annual Precip. (mm)	350	400	450
Topography	Strongly rolling	Strongly hummocky	hummocky
Elevation range (m)	44	12	5
Parent Material	till	fluvial/till	till/fluvial
Soils	Dark Brown Chernozemic, Regosolic, Gleysolic	Thin Black Chernozemic, Gleysolic, Regosolic	Black Chernozemic, Solonetzic, Gleysolic

Commercial continuous yield monitors were installed on the four different types of farm combines. The yield monitors were interfaced with the portable GPS receivers when each of the project fields were harvested. Variable rate fertilization was done with a pneumatic banding applicator with two tanks for individual control of the rate of nitrogen (urea) and phosphate (mono ammonium phosphate) fertilizer. Research in the Canadian prairies has shown this to be the most efficient and environmentally benign method of fertilizer application. The delivery rate from the two tanks of fertilizer could be adjusted instantly, according to a prescription map, as the tractor-applicator moved across the field.

The GPS receivers used were 10 channel C/A code narrow correlator spacing NovAtel 951 GPSCard connected to NovAtel Model 501 antennas (Fenton et al. 1991, Van Dierendonck et al. 1992). Differential corrections were done in post processing of harvest operations. Three dimensional accuracy at the sub-decimeter level were achieved. For variable rate fertilizer application the base station set up in the field corner and the mobile receiver on the tractor were linked with two short range wireless radio modems in the low 900 MHz RF range. This provided DGPS positions in real-time. The integrated GPS system and it's performance is described by Lachapelle et al. (1994).

For the parent project 18 to 26 hectares at each location was soil sampled using a 68 by 68 m grid in the fall of each year. At each grid-line intersection, composite samples of 12 to 15 cores were taken at four increments to a depth of 90 cm. The soil test results were used in conjunction with yield maps, aerial photographs and topographic maps to construct fertilizer application maps. Map units with similar levels of N and P were defined. The optimum application rates for N and P were estimated based on soil test values, yield and landscape features. The Bow Island and Mundare sites were mapped for salinity using an EM38 salinity meter and GPS as described by Cannon et al. (1994).

Fertilizer rate experiments were conducted in strips 13 m wide across the full length of the fields. Four constant rates of N with one rate of P were used in one block. The other block had three constant rates of P with one rate of N. Each rate was replicated twice. On an adjacent area, variable rate applications (N and P) were compared to constant rate applications in alternating strips.

The public domain, raster based GIS, GRASS (Geographic Resource Analysis Support System) was used for mapping, overlays and data analyses (U.S. Army, 1993). The landform regimes were described using the system of Pennock et al. (1987).

In 1994 and 1995 site specific sampling points were characterized at all sites for the purposes of providing input for the EPIC model. Holes were dug at the shoulder, backslope and footslope position of four hills at each site. Soil profiles were described and sampled for physical and chemical analyses. The position of each hole was established with DGPS.

A GRASS statistical routine was developed (s.rstats) to search the raw yield data around the soil description hole for yield data from the combine harvestor. The routine will for a desired number of yield data points (e.g. n=15) around a point providing the mean and standard deviation of yield as well as the closest value to the point and its distance along with the mean distance of all points. This allows the GIS operator to determine and control the areal extent of the yield data to conform to landscape position, fertilizer treatment or other field features. The resultant yields and their variances are used to assess the performance of the EPIC crop model for site specific management.

The EPIC model version 3090 was used on a DOS platform. Detailed profile data were used to generate EPIC input files. Although some climate data was recorded at each field site, the simulations reported here are using the nearest weather station data (less than 30 km distance). Daily data from the nearest stations was obtained and assembled into the appropriate format for EPIC. The crop at two of the sites in 1994 (Hussar, Stettler) was spring wheat and canola at the other (Mundare). The plant growth parameters for the Canadian prairies from Kiniry et al. (1995) were used instead of the defaults.

RESULTS and DISCUSSION

Only results from the three dryland farms will be presented and discussed. Delays in obtaining the yield data for 1995 have prevented presentation of that data at this time.

The performance of the DGPS receivers and the software for solution of position allowed for vertical accuracies in the range of 70 mm and are futher reported elsewhere (Lachapelle et al., 1994). The amount of elevation data collected in each field was dependent upon the duration of harvest operations but was usually not less than 50,000 observations for an 80 hectare field. The consequence of using high precision DGPS each year, or several times per year (e.g. planting, harvesting, spraying) is that very large data sets can be accumulated in a very short period of time as a by-product of other precision farming operations. The data was not geometrically even as the GPS receiver was mounted on the combine harvester which cut the crop at six to nine meter widths depending upon the crop and combine design. The data collection rate was constant (1 Hz) and since combines have hydrostatic drives it allows a continuously variable forward speed which usually had data spaced from one to 2.5 m within the lines. This presented some issues of either de-densifying data or using robust interpolation techniques.

Satisfactory topographic contours were developed in GRASS using the thin plate spline method with tension and smoothing ("s.surf.tps"). Linearity of the raw data was reflected in the result of using an inverse distance weighted routine ("r.idw"). If several data sets were to be combined as mentioned above, the data density would become so great that any interpolation method would likely be adequate.

The characteristics of the soils varied greatly with landscape position (Table 2). Often the difference in classification at each position was at the soil order level of the Canadian System of Soil Classification.

Yields measured by the combine harvesters were found to vary by slope position at all sites. An example of one site and year is presented in Table 3 of data from the whole field when segmented according to landscape element (after Pennock et al. 1987).

The measured yields by landscape position at the four transects in each field also displayed the same trend however, a difference with respect to the lower slope position is evident between the site with conventional tillage (Stettler) and the site which has been under a direct seeding system for 10 years (Hussar) (Figure 1). The effect of the thick straw mulch on temperature and moisture as well as the finer soil textures have all contributed to less of a difference between slope position. The lower slope position at Hussar was very moist and crop lodging did occur to further decrease yield. The wheat grown at Hussar in 1994 was a newly licensed variety of a new wheat type (Canada Prairie Spring wheat, AC Taber variety). EPIC crop parameters may not be appropriate for that type of wheat.

EPIC yields were consistently lower for the three sites in 1994 (1.4 Mg/ha and 2.0 Mg/ha respectively). Likewise, the coefficient of variation for EPIC predicted yields was nearly half the CV of the measured yields (CV for predicted was 27% and for actual, 49%). Linear regression analysis indicated a significant, but weak, relationship between the predicted and measured yields (Figure 2). Inspection of the canola crop and wheat crops independently showed no better agreement of one over the other. Recent calibration and validation work with EPIC in France showed a coefficient of determination of half the value obtained here (Cabelguenne et al. 1990). They used nearly same number of data points and the crop was only wheat. Wheat was found to have the lowest coefficients of determinations compared to other crops such as soybeans and sorghum (Cabelguenne et al. 1990). Improved relationships are expected with daily weather data from an on site station as well as mass transfer of runoff from one landscape element to the other.

In order for integrated crop models to be accepted, the ability to fairly represent yield potential on a site specific basis is needed. The benefit of using an integrated model is that other features of the model can then be used with little additional effort. This latter feature may be very appealing for wide spread use in the agriculture industry. Models for pesticide leaching, erosion and tillage are more likely to be used if they come as part of an agronomic model that can be used for prescription mapping and risk assessment.

Table 2. Soil characteristics by landscape position for three dryland sites.

		Landscape Position					
		Footslope		**Backslope**		**Shoulder**	
Site		Mean	STD	Mean	STD	Mean	STD

Hussar	OM%	3.2	0.3	3.9	2.8	3.3	2.3
	Depth to B horizon (cm)	11	1	20	9	10	1
	pH	6.4	0.3	6.7	0.9	6.6	0.9
	Sand%	37	8	40	6	38	6
	Clay%	29	6	26	4	28	5
	Nitrate (ppm)	10	1	16	15	17	19
	Phosphate (ppm)	13	4	12	8	9	5
Mundare	OM%	3.6	1.2	3.6	1.7	3.3	1.5
	Depth to B horizon (cm)	44	26	30	3	16	7
	pH	5.8	0.2	7.0	1.1	6.1	0.7
	Sand%	43	10	50	8	49	7
	Clay%	20	3	20	3	21	2
	Nitrate (ppm)	20	2	16	8	12	7
	Phosphate (ppm)	22	11	16	8	14	8
Stettler	OM%	4.5	0.1	2.5	0.3	2.5	0.8
	Depth to B horizon (cm)	58	23	25	13	13	3
	pH	6.4	0.7	6.5	0.2	7.0	0.5
	Sand%	61	1	70	6	71	9
	Clay%	15	0	13	3	14	4
	Nitrate (ppm)	25	1	12	7	11	3
	Phosphate (ppm)	10	5	13	8	13	9

Table 3. Yield of fertilized and unfertilized spring wheat by landscape position at the Hussar site, 1994.

	Spring wheat yield (Mg/ha)		
	Landscape position		
Fertilizer Treatment	Shoulder	Backslope	Footslope
Fertilized (F)	3.59	3.74	3.67
Unfertilized (U)	2.54	2.64	3.32
F - U	1.05	1.11	0.35

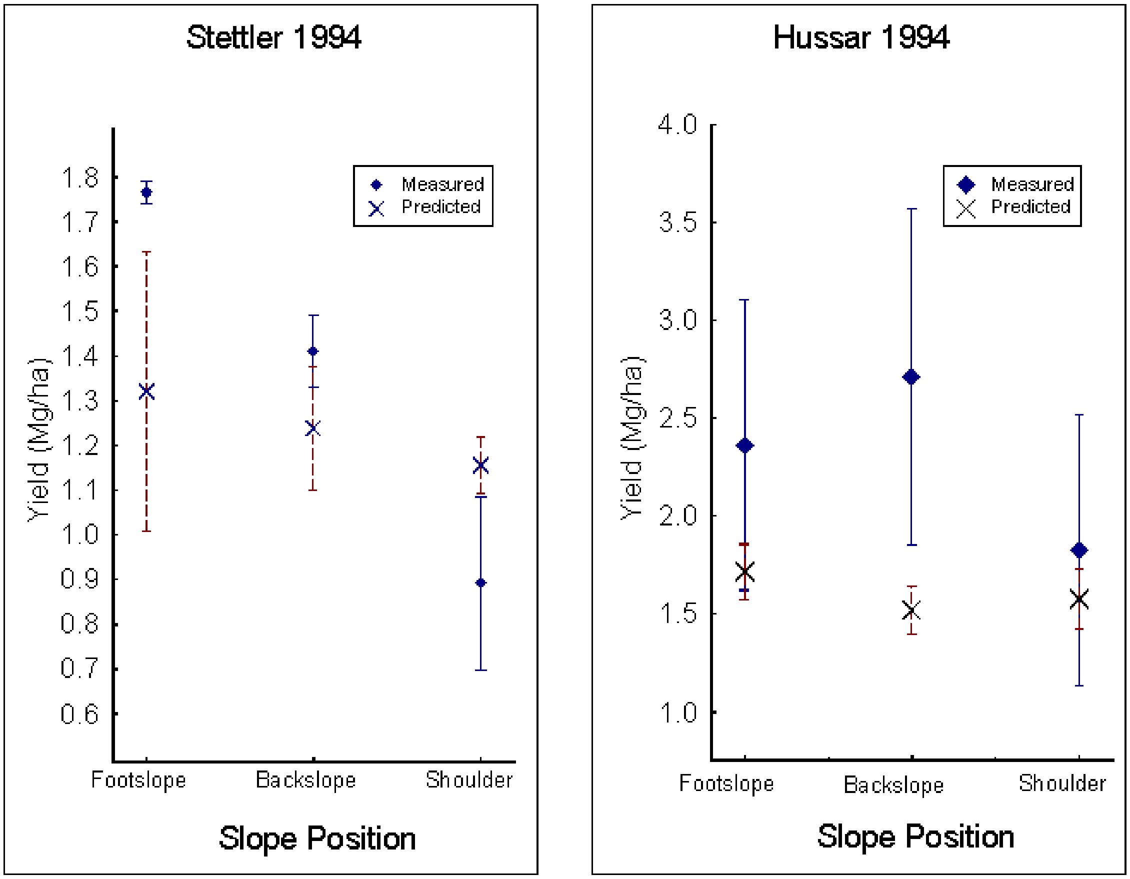


Figure 1. Measured and predicted yields at two dryland sites.

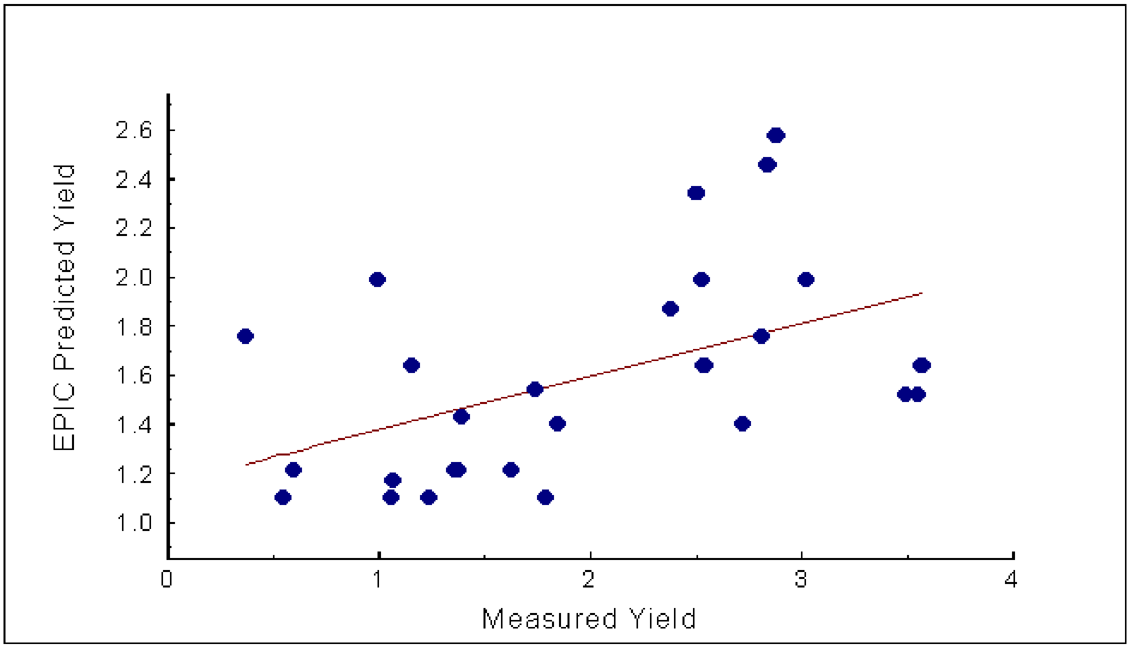


Figure 2. Regressed EPIC yields for three dryland sites in 1994 (wheat and canola).

Regression ANOVA Table for Figure 2.

Source	Sum of sq.	Deg of Free.	Mean Sq.	F-Ratio	Prob>F
Model	1.13	1	1.13	8.31	0.01
Error	3.59	28	0.14		
Total	4.72				

Coefficient of determination 0.24

Variable	Coefficient	S.E.E.	T-Statistic	Prob>t
Constant	1.16	0.17	6.87	0.00
Yield	0.22	0.08	2.81	0.01

ACKNOWLEDGMENTS

Funding for this project was provided by the Canada-Alberta Environmentally Sustainable Agriculture (CAESA) agreement. The project benefited from the field and computer skills of Germar Lohstraeter and Sheilah Nolan of Alberta Agriculture, Food and Rural Development. GPS support was provided by the University of Calgary, Department of Geomatics.

REFERENCES

Cabelguenne, M., Jones, C.A., Marty, J.R., Dyke, P.T. and Williams, J.R. (1990) Calibration and validation of EPIC for crop rotations in southern France. Agric. Systems 33: 153-171.

Fenton, P, Falkenberg, W.,Ford, T., Ng,K., Van Dierendonck, A.J. (1991) NovAtel's GPS receiver, the high performance OEM sensor of the future. Proc. GPS91, The Institute of Navigation, Alexandria, VA, pp. 49-58.

Goddard, T.W., Lachapelle, G., Cannon, M.E., Penney, D.C., and McKenzie, R.C. (1995) The potential of GPS and GIS in precision agriculture. Proc. Geomatique V: "La Route De L'Innovation". November 9-10, Montreal, P.Q., Canada

Kiniry, J.R., Major, D.J., Izaurralde, R.C., Williams, J.R., Gassman, P.W., Morrison, M., Bergentine, R. and Zentner, R.P. (1995) EPIC model parameters for cereal, oilseed, and forage crops in the norther Great Plains region. Can. J. Plant Sci. 75: 679-688.

Lachapelle, G., Cannon, M.E., Gehue, H., Goddard, T. and Penney, D. (1994) GPS system integration and field approaches in precision farming. Navigation 41(3): 323-335.

Moulin, A. P. and Beckie, H.J. (1993) Evaluation of the CERES and EPIC models for predicting spring wheat grain yield over time. Can. J. Plant Sci. 73: 713-719.

Pennock, D.J., Zebarth, B.J., and De Jong, E. (1987) Landform classification and soil distribution in hummocky terrain, Saskatoon, Canada. Geoderma 40: 297-315.

Toure, A., Major, D.J. and Lindwall, C.W. (1995) Sensitivity of four wheat simulation models to climate change. Can. J. Plant Sci. 75: 69-74.

Van Dierendonck, A.J., Fenton, P. and Ford, T. (1992) Theory and performance of narrow correlator spacing in a GPS receiver. The Institute of Navigation, Alexandria, VA, Vol. 39, No. 3, pp. 265-283.

Tom Goddard, Len Kryzanowski, Karen Cannon
Alberta Agriculture, Food and Rural Development
#206, 7000 - 113 St.
Edmonton, AB Canada
T6H 5T6
phone: 403-422-6530
fax: 403-422-0474
goddard@gpu.srv.ualberta.ca
kryzano@agric.gov.ab.ca

Cesar Izaurralde, Tim Martin
Dept. Renewable Resources
751, GSB
University of Alberta
Edmonton, AB Canada
T6G 2E1
phone: 403-492-5104
fax: 403-492-4323

Cesar.Izaurralde@ualberta.ca
Tim.Martin@ ualberta.ca

Kumar C.S Navulur, Bernard A. Engel

Predicting Spatial Distributions of Vulnerability of Indiana State Aquifer Systems to Nitrate Leaching using a GIS

ABSTRACT

Regional scale analysis identifying the problem areas of nitrates leaching from agricultural management systems will aid in efficient groundwater management strategies. Preliminary screening analysis to evaluate the vulnerability of groundwater systems of Indiana to nitrate pollution was carried out using a modified DRASTIC and SEEPAGE analysis at 1:250,000 scale. The state soils geographic database (STATSGO) was used to extract the soil information required for the analysis. The accuracy of the results from the above analysis was statistically evaluated by comparing the results with groundwater quality data sampled across the state. The comparison showed a correlation coefficient of 0.67 and showed that these regional scale analyses show a great deal of potential as screening tools for policy decision making in groundwater management.

Keywords: STATSGO, DRASTIC, SEEPAGE

INTRODUCTION

Groundwater contamination by nitrates due to application of fertilizers and livestock waste in agricultural management systems is of wide concern. Reported findings of groundwater contamination in wells in New York led the US Environmental Protection Agency (USEPA) to conduct a nation wide survey on well contamination in the United States in 1989. Samples taken from a total of 135 samples from 564 community wells and 783 rural drinking water wells were tested for presence of nitrates, pesticides, and pesticide breakdown products (USEPA, 1990). The wells selected, statistically, are representative of more than 94,600 wells in approximately 38,300 community water systems and more than 10.5 million rural wells. Over 52% of the community water systems and 57% of the rural domestic wells tested contained nitrates (USEPA,1992).

Indiana has abundant groundwater systems providing drinking water for 60 percent of its population. In a study on well water quality, 4% of wells tested in Indiana had detectable pesticides. Also 10% of private wells and 2% of non-community wells contained excessive nitrate levels above the MCL of 40 parts per million (Indiana Department of Environmental Management, 1989). The protection of these drinking water systems from nitrate contamination is of great importance.

Statewide maps showing the areas vulnerable to groundwater contamination would have many potential uses such as implementation of groundwater management strategies to prevent degradation of groundwater quality and monitoring the groundwater systems. These maps would be helpful in evaluating existing and potential policies for groundwater protection. Groundwater models such as SEEPAGE and DRASTIC can be applied on a regional scale to develop such maps. The data layers required for these models are commonly available data such as pH, organic matter content, etc. For most states, the statewide groundwater vulnerability maps using DRASTIC were produced from 1:2,000,000 scale data (Aller et al., 1987). The USEPA (1992) found that these maps did not correlate well with the water quality analysis performed for the national survey of pesticides in drinking water wells. More detailed and accurate maps are needed by states to implement groundwater management programs. The state soils geographic (STATSGO) database at the 1:250,000 scale might be useful for studies at a larger scale.

While regional scale models can be developed using the commonly available data, field scale models require detailed inputs to model contaminant transport in the root zone and are helpful in investigating areas of high vulnerability from the regional scale maps. These will also be helpful in suggesting the conservation practices that can mitigate the pollution. GLEAMS, NLEAP, LEACHMN, and CMLS are such models that can consider detailed inputs like evaporation and management practices for estimating the leaching potential of the nitrates. These models can be applied to study areas of disagreement in the regional scale vulnerability maps with water quality data samples.

The Geographic Information Systems (GIS) environment is being widely applied for diverse applications in resources management and other areas. It offers the facilities to store, manipulate and analyze data in different formats and at different scales (Evans, 1990). Integration of groundwater quality assessment models in a GIS will allow use of these models for different scenarios (management practices, land uses, etc.).

OBJECTIVES

The primary objectives of this research are:

1. Evaluate Indiana's groundwater vulnerability to nitrate pollution potential using the DRASTIC and SEEPAGE models.
2. Determine the spatial auto-correlation of nitrate detections in groundwater across the state to determine whether point or non-point sources are likely the cause of such pollution.
3. Validate the accuracy of the approach by comparing the vulnerability maps with existing well water quality data sampled across the state.

LITERATURE REVIEW

DRASTIC

DRASTIC is a groundwater quality index for evaluating the pollution potential of large areas using the hydrogeologic settings of the region (Aller et al., 1985, Aller et al., 1987, Deichert et al., 1992). This model was developed by EPA in the 1980's. DRASTIC includes various hydrogeologic settings which influence the pollution potential of a region. A hydrogeologic setting is defined as a mappable unit with common hydrogeologic characteristics. This model employs a numerical ranking system that assigns relative weights to various parameters that help in the evaluation of relative groundwater vulnerability to contamination. The hydrogeologic settings which make up the acronym DRASTIC are:

[D] Depth to water table: Shallow water tables pose a greater chance for the contaminant to reach the groundwater surface as opposed to deep water tables.

[R] Recharge (Net): Net recharge is the amount of water per unit area of the soil that percolates to the aquifer. This is the principal vehicle that transports the contaminant to the groundwater. The more the recharge, the greater the chances of the contaminant to be transported to the groundwater table.

[A] Aquifer Media: The material of the aquifer determines the mobility of the contaminant through it. An increase in the time of travel of the pollutant through the aquifer results in more attenuation of the contaminant.

[S] Soil Media: Soil media is the uppermost portion of the unsaturated / vadose zone characterized by significant biological activity. This along with the aquifer media decides the amount of percolating water to the groundwater surface. Soils with clays and silts have larger water holding capacity and thus increase the travel time of the contaminant through the root zone.

[T] Topography (Slope): The higher the slope, the less is the pollution potential due to higher runoff and erosion rates which include the pollutants that infiltrate into the soil.

[I] Impact of Vadose Zone: The unsaturated zone above the water table is referred to as the vadose zone. The texture of the vadose zone determines the time of travel of the contaminant through it. Authors of this model suggest that the layer that most restricts the flow of water be used.

[C] Conductivity (Hydraulic): Hydraulic conductivity of the soil media determines amount of water percolating to the groundwater through the aquifer. For highly permeable soils, the travel time of pollutant is decreased within the aquifer.

The major assumptions outlined in DRASTIC are:

1. The contaminant is introduced at the surface
2. The contaminant reaches groundwater by precipitation
3. The contaminant has the mobility of water
4. The area of the study site is greater than 100 acres

DRASTIC evaluates pollution potential based on the above seven hydrogeologic settings. Each factor is assigned a weight based on its relative significance in affecting pollution potential. Each factor is further assigned a rating for different ranges of the values. The typical ratings range from 1-10 and the weights from 1-5. The DRASTIC Index, a measure of the pollution potential, is computed by summation of the products of rating and weights of each factor as follows:

DRASTIC Index = DrDw + RrRw + ArAw + SrSw + TrTw + IrIw + CrCw

Where

Dr = Ratings to the depth to water table
Dw = Weights assigned to the depth to water table
Rr = Ratings for ranges of aquifer recharge
Rw = Weights for the aquifer recharge
Ar = Ratings assigned to aquifer media
Aw = Weights assigned to aquifer media
Sr = Ratings for the soil media
Sw = Weights for soil media
Tr = Ratings for topography (slope)
Tw = Weights assigned to topography
Ir = Ratings assigned to vadose zone
Iw = Weights assigned to vadose zone
Cr = Ratings for rates of hydraulic conductivity
Cw = Weights given to hydraulic conductivity

The higher the DRASTIC index, the greater the relative pollution potential. The DRASTIC index can be further divided into four categories: low, moderate, high, and very high. The sites with high and very high categories are more vulnerable to contamination and hence can be reviewed by a specialist. These weights are relative and a site with low pollution potential need not necessarily mean that it is free from groundwater contamination but it is relatively less susceptible to contamination compared to the sites with high or very high DRASTIC ratings.

The USEPA (1992) analyzed the results of the National Survey of Pesticides in Drinking Water Wells data and the qualitative DRASTIC scores developed by Aller et al. (1987) at a 1:2,000,000 scale. County level DRASTIC scores (an aggregated score for each county) and subscores were computed and 90 counties were selected for analysis with nitrate data from wells sampled in the study. The results showed that DRASTIC performed very poorly for the selected counties. Hence for implementation of groundwater quality management plans at a regional scale more detailed vulnerability map is needed. The proposed study will use more detailed data at the 1:250,000 scale for estimating DRASTIC indices and additional data to improve vulnerability estimates.

SEEPAGE

The System for Early Evaluation of Pollution potential of Agricultural Groundwater Environments (SEEPAGE) model is a combination of three models that was adapted to meet SCS (Soil Conservation Service, recently renamed the Natural Resources Conservation Service) needs to assist field personnel (Moore et al., 1990, Richert et al. 1992, Engel et al. 1992). SEEPAGE considers various hydrogeologic settings and physical properties of the soil that affect groundwater vulnerability to pollution potential. This is also a numerical ranking model that considers contamination from both concentrated and dispersed sources.The SEEPAGE model considers the following parameters:

1. Soil slope
2. Depth to water table
3. Vadose zone material
4. Aquifer material
5. Soil depth
6. Attenuation potential

The attenuation potential further considers the following factors:

1. Texture of surface soil
2. Texture of sub soil
3. Surface layer pH
4. Organic matter content of the surface
5. Soil drainage class
6. Soil permeability (Least permeable layer)

Each factor is assigned a numerical weight ranging from 1-50 based on its relative significance, with the most significant parameter affecting the water quality assigned a weight of 50 and the least significant assigned a weight of 1. The weights are different for concentrated sources (site specific), and dispersed sources (non specific sources). Similar to DRASTIC, each of the factors can be divided into ranges and ratings assigned varying from 1-50. The ratings of the aquifer media and vadose zone are subjective and can be changed for a particular region. Once the scores of the six factors are obtained, these are summed to get the SEEPAGE Index Number (SIN). These values are representative of the pollution potential where a high SIN value implies relatively more vulnerability of the groundwater system to contamination. The SIN numbers are arranged into four categories of pollution potential: low, moderate, high, and very high. A high or very high SIN category indicates that the site has significant constraints for groundwater quality management (Richert et al., 1992).

Engel (1992) used GRASS to carry out SEEPAGE analysis to evaluate the pollution potential of groundwater systems for the Kennedy Space Center, Florida. The data layers for carrying out the analyses were integrated in the GRASS GIS environment. In the research reported in this paper, SEEPAGE is used along with the DRASTIC model to carry out the regional scale studies. Both analyses were carried out in the ARC/Info GIS environment.

METHODOLOGY

Developing the Data Layers

The STATSGO database from SCS comes at a scale of 1:250,000 and is distributed in different data formats. This database in the ARC/Info format was used in this study. The database is organized into map units which have as many as 21 components. These map components have information assigned to layers of soil horizons. Each of the layers are attributed various soil properties such as pH, organic matter content, etc. (SCS, 1992). Each of the properties is assigned a high and a low value for a mapunit. The STATSGO map for Indiana is available in the vector format and was used as the base map for the DRASTIC and SEEPAGE analyses. The hydrogeologic parameters required for the analysis were identified from the corresponding Info data tables and map layers for each of the layers were created using database management tools in ARC/Info. Codes were developed in Arc Macro Language (AML) to automate the process of extracting the information from the database files and assigning the corresponding ratings required for the analyses.

The water table depth data layer was created by interpolation from a set of approximately 7200 data points of water table depth in well sites distributed across the state. Aquifer media and vadose media were extracted from the glacial geology map of Indiana. The aquifer recharge was computed using the percolation index (PI) in the NLEAP model (Deichert. et al., 1992, Follet et al., 1991, Follet et al., 1994). The hydrologic soil groups required for computing aquifer recharge were extracted from the STATSGO database and seasonal and annual precipitation distribution across Indiana was computed using forty year precipitation records from twenty weather stations. The land use data was obtained from the SCS and the fertilizer data extracted from agricultural statistics. These data layers, along with the hydrogeologic settings, were employed in this analysis.

Carrying out the Analyses

The data layers of the hydrogeologic settings were assigned the corresponding ratings [tables 1 and 2] and were converted from vector to raster layers in ARC/Info (ESRI 1992a, ESRI 1992b, and ESRI 1992c). A Graphical User Interface was developed using the form menus in ARC/Info for carrying out the analyses. Using the GUI the DRASTIC and SEEPAGE analyses were carried out. The DRASTIC and SEEPAGE analyses were modified by considering the additional data layers land use and fertilizer usage. The agricultural statistics were used to extract the crop information (type of crop, yield, harvest date, etc.), and the fertilizer applied on a county scale. Based on the crop N uptake and N content of fertilizer, the excess fertilizer applied on farm land was computed for the state. The DRASTIC Index map and SEEPAGE Index Number (SIN) map were reclassified into four categories: [tables 1 and 2] low, moderate, high and very high.

DRASTIC Index	Low	Moderate	High	Very High
Values	1-140	141-180	181-230	> 230

Table 1: Reclass Table for Modified DRASTIC Ratings

SEEPAGE Index	**Low**	**Moderate**	**High**	**Very High**
Values	1-89	90-144	145-209	> 210

Table 2: Reclass Table for Modified SEEPAGE Ratings

Spatial Correlation Studies to Eliminate Point Source Pollutants

Spatial auto-correlation refers to the spatial ordering of a single variable and to the relationship between pairs of observations of this variable. The ordering of n observed values of some variable X is usually described with the aid of a connectivity matrix, C. Non-zero c_{ij} entries in the n x n matrix indicate that the corresponding polygons are juxtaposed. For data measured on a interval / ratio scale the statistics geary ratio and moran coefficient can be used (Griffith, 1987).

Geary Ratio

The Geary Ratio is an index for interval / ratio data that is based upon paired comparisons of juxtaposed map values. It may be calculated as :

$$GR = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(x_i - x_j)^2}{2\left(\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\right)\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

The meaning of this index is fairly straight forward: as similar values tend to clump together ($c_{ij} = 1$), then the Geary ratio approaches to zero. If dissimilar values tend to clump together, the geary ratio approaches 2.

Moran Coefficient

Another index for interval / ratio based data is Moran Coefficient and may be calculated as :

$$MC = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

The expected value of MC is (-1/n-1). For similar values in juxtaposition, MC-> 1 and for dissimilar values in juxtaposition, MC -> -1. As the Geary ratio deals with paired comparisons, the Moran coefficient deals with covariations.

As the DRASTIC and SEEPAGE analysis predict vulnerability from non-point source pollution, it was desirable to eliminate detections due to point source pollutions. The water quality data containing the nitrate detections in well sites was imported as a point coverage into ARC/Info and Thiessen polygons were created for the point coverage. The spatial auto-correlation statistics, Moran Coefficient and Geary Ratio were computed for the datasets. The Moran Coefficient value was computed as 0.79 and Geary Ratio as

-0.06 which indicated that a spatial correlation exists among the detects. It is assumed that computation of these spatial statistics for different combinations of nitrate detections will help in detecting values that significantly affect the Geary Ratio and Moran Coefficient, which might be an indication of point source pollution.

Validation of the Accuracy of the Regional Scale Vulnerability Maps

The modified DRASTIC and SEEPAGE results were compared with nitrate detections in 380 well sites sampled across Indiana. The nitrate detections were categorized into four categories: Low 0-5 ppm; Moderate 5-15 ppm; High 15-30 ppm; Very high > 30 ppm. As the well water quality samples for nitrate detections were not uniformly distributed across the state, the modified DRASTIC and SEEPAGE vulnerability ratings at the corresponding sites were extracted and compared with the nitrate detections in the well water quality database (Dou and Woldt, 1994). The results from the modified DRASTIC and SEEPAGE analysis were compared with the results from conventional DRASTIC and SEEPAGE analyses [tables 3, 4, and 5].

RESULTS

The DRASTIC and SEEPAGE analyses were carried out in ARC/Info, using the GUI, to create the GIS layers shown in figures 1 and 2. The modified techniques considering the additional data layers land use and fertilizer usage were carried out in the grid sub module. The results from the conventional DRASTIC model indicate that 58% of the groundwater systems in Indiana fall under the moderate vulnerability category and 23% under high and very high pollution potential. There was a 24% increase in the areas categorized as low vulnerability using the modified DRASTIC approach. The conventional SEEPAGE approach predicted around 75% of the state having moderate vulnerability and considering the additional datalayers land use and fertilizer usage did not change the predictions. Both models predicted 50% or more of Indiana aquifer systems to have moderate vulnerability ratings.

The results from the conventional and modified DRASTIC and SEEPAGE approaches were compared with the database containing observed nitrate detections in wells. The conventional DRASTIC and SEEPAGE models predicted 80% of the high and very high vulnerable areas correctly. There was a 20% increase in accuracy in predicting low vulnerability areas using the modified DRASTIC technique [table 4]. The consideration of additional data layers in SEEPAGE did not improve the accuracy of predictions [table 5].

	DRASTIC Ratings			
Observed	**Low**	**Moderate**	**High**	**Very High**
Low	25	146	138	0
Moderate	1	17	63	0
High	0	2	18	2
Very High	0	2	2	1

Table 3: Comparison of Conventional DRASTIC Ratings With Observed Nitrate Detections

	Modified DRASTIC Ratings			
Observed	**Low**	**Moderate**	**High**	**Very High**
Low	86	85	138	0
Moderate	7	11	63	0
High	0	1	20	1
Very High	0	2	2	1

Table 4: Comparison of Modified DRASTIC Ratings With Observed Nitrate Detections

	SEEPAGE Ratings			
Observed	**Low**	**Moderate**	**High**	**Very High**
Low	9	183	117	0
Moderate	2	22	57	0
High	0	2	20	0
Very High	0	2	3	0

Table 5: Comparison of Modified SEEPAGE Ratings With Observed Nitrate Detections

The Pearson Correlation Coefficient (r) was computed as 0.61 between conventional DRASTIC predictions to the observed nitrate detections. The addition of data layers land use and fertilizer application improved the r value to 0.67. DRASTIC and SEEPAGE ratings were generally conservative in prediction. The areas predicted to have low vulnerability but which had observations of high or very high nitrate concentrations are of concern and are being investigated further. Such high observations of nitrates may be the result of point sources. Using SAS (Statistical Analysis Software) was used to carry out regression analysis on all the possible combinations of DRASTIC factors including the additional data layers to determine the best possible combination that predicts the nitrate detections accurately. The aquifer recharge data layer was found to have the least significant effect on nitrate detections and could be dropped from the model. This is largely due to the relatively small difference of this parameter within the study area. Studies are also underway to perform sensitivity analysis on the weights assigned to each of the factors.

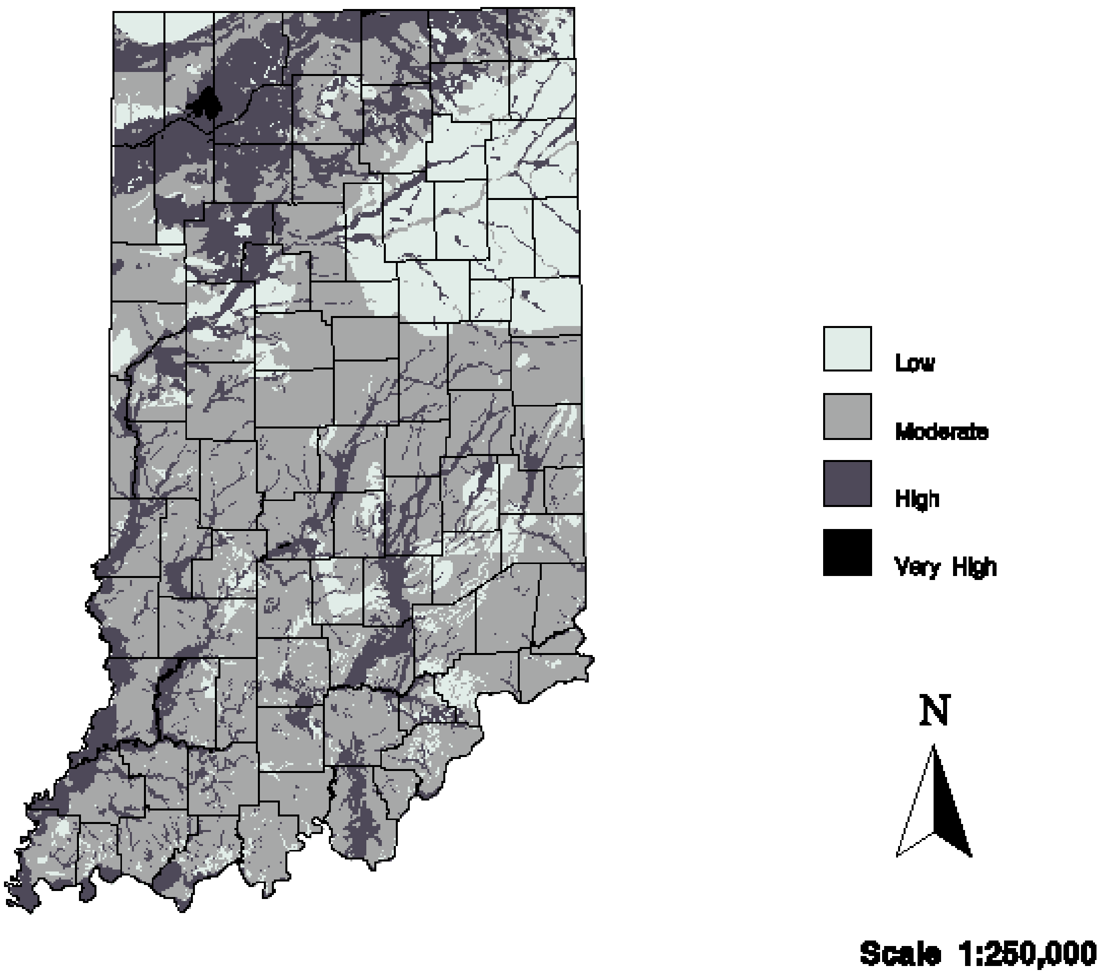


Figure 1: Groundwater Vulnerability Map to Nitrate Pollution Using Modified DRASTIC Technique

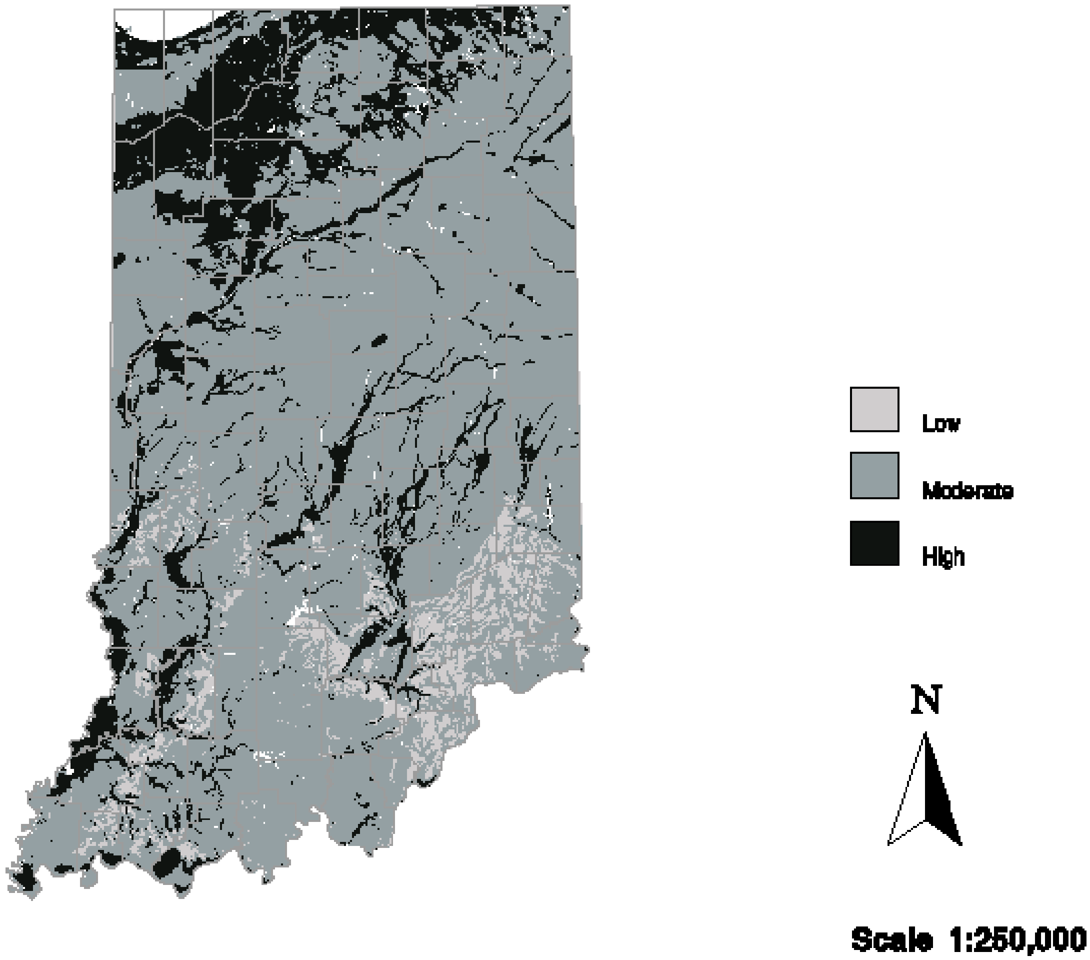


Figure 2: Groundwater Vulnerability Map to Nitrate Pollution Using Modified SEEPAGE Technique

SUMMARY

Regional scale groundwater vulnerability studies to nitrate pollution from non-point sources using DRASTIC and SEEPAGE analyses were conducted at the 1:250,000 scale for Indiana. These models were integrated with the ARC/Info GIS environment. DRASTIC and SEEPAGE analyses categorized most of the Indiana aquifer systems as moderately vulnerable to pollution. Additional data layers, land use and fertilizer usage, were considered in DRASTIC and SEEPAGE in an attempt to improve the results. The results from conventional and modified DRASTIC and SEEPAGE analyses were compared with the nitrate detections in well water quality samples across the states. The results indicated that the models performed well in predicting the sites with high and very high nitrate detections.

Statistics, Geary ratio and Moran coefficient were computed to assess the spatial auto-correlation of nitrate detections. The results indicated that spatial correlation exits between the occurrences of nitrate pollution. This suggests that regional-scale factors and processes are of importance and indicating that non-point sources of nitrates are important to the quality of groundwater.

The modified DRASTIC and SEEPAGE analysis show a great deal of potential as screening tools for policy decision making in groundwater management. These analyses should not be used to replace detailed studies but should be applied to screen areas of high and very high vulnerability that a site-specialist should concentrate on. These analyses can be further modified by including additional data showing the excess amount of nitrate applied, crop rotations, and livestock nutrient value applied to farms. It is anticipated that this will improve the accuracy of the vulnerability estimates.

GIS are an effective tool for carrying out groundwater vulnerability studies. The integration of the groundwater quality models within the GIS environment results in easy implementation of the analyses for different agricultural management practices. Automated data extraction to develop data layers required for the DRASTIC and SEEPAGE analysis from the STATSGO database will enable analysis for the whole Unites States at a 1:250,000 scale. When more data becomes available, the field scale simulations of more detailed models like NLEAP can be used for regional scale studies. It is anticipated that the studies on the spatial and scale variability of inputs on field scale models will lead to specific criteria with regards to level of detail of inputs required for conducting regional scale analyses.

REFERENCES

Aller, L., Bennett, T.,Lehr, J. H., and Petty, R.J., 1985, "DRASTIC : A standardized system for evaluating groundwater pollution potential using hydrogeologic settings," U.S. EPA, Robert S. Kerr Environmental Research Laboratory, Ada, OK, EPA/600/2-85/0108, 163pp.

Aller, L., Bennet, T., Lehr, J. H., Petty, R. J., and Hackett, G., 1987, "DRASTIC : A standardized system for evaluating groundwater pollution potential using hydrogeologic settings," EPA-600/2-87-035.

Deichert, L. A. and Hamlet, J. M., 1992, "Non-point groundwater pollution potential in Pennsylvania," ASAE International Winter Meeting, Nashville, Tennessee, 15-18 December, 1992, Paper No. 922531.

Dou, C and Woldt., D. E., 1994, "Application of geostatistics for mapping nitrate contaminated groundwater," ASAE International Summer Meeting, Kansas City, Missouri, 20-23 June, 1994, Paper No. 942095.

DRASTIC User Manual, 1987, Robert S. Kerr Environmental Laboratory, U.S. Environmental Protection Agency, ADA, OK.

Engel, B. A., 1992, "Water quality modeling using Geographic Information System

(GIS) data," 1992 NASA/ASEE summer faculty fellowship program, John F. Kennedy Space Center, University of Central Florida, Aug. 10, 1992.

ESRI, 1992a, "ARC/Info user's guide," Cell based modeling with grid. March, 1992.

ESRI, 1992b, "ARC/Info user's guide," AML user's guide. May, 1992.

ESRI, 1992c, "ARC/Info command references," Arc command references. May, 1992.

Evans, Barry M., 1990, "GIS in groundwater systems management," Journal of Soil and Water Conservation, Vol. 45:246-8.

Follet, R.F., Keeny, D. R., and Cruse, R. M., 1991, "Managing nitrogen for groundwater quality and farm profitability," SSSA, Madison, WI.,pp. 285-322.

Follet, R. F., Shaffer, M. J., Brodahl, M. K., and Reichman, G. A., 1994, "NLEAP simulation of residual soil nitrate for irrigated and non- irrigated corn," Journal of Soil and Water Conservation, Vol. 49(4):375-82.

Griffith, D. A., 1987, "Spatial auto-correlation - A primer", Association of American Geographers.

Indiana Department of Environmental Management, 1989, "Groundwater section," Indiana groundwater protection, A guide book. June 1989.

Moore, and John S., 1990, "SEEPAGE: A system for early evaluation of the pollution potential of agricultural groundwater environments," USDA. SCS, Northeast Technical Center, Geology Technical Note 5.

Richert, S. E., Young, S. E., and Johnson, C, 1992, "SEEPAGE: A GIS model for groundwater pollution potential," ASAE 1992 International Winter Meeting, Nashville, Tennessee, 15-18 December, 1992, Paper No. 922592.

SCS, 1992, "State Soils Geographic Database (STATSGO) User's Guide, 1992," SCS, Publication No. 1492, August 1992.

USEPA, 1990, "National Pesticide Survey: Phase I Report," USEPA Washington DC.

USEPA, 1992, "Another Look: National Survey of Pesticides in Drinking Water Wells Phase II Report," USEPA 579/09-91-020. EPA Washington, DC.

Kumar C.S Navulur, Graduate Research Assistant, Department of Agricultural and Biological Engineering, Purdue University, W. Lafayette, IN 47907-1146, ph 317-494-1196, email: navulur@ecn.purdue.edu

Bernard A. Engel, Associate Professor, Department of Agricultural and Biological Engineering, Purdue University, W. Lafayette, IN 47907-1146, ph 317-494-1198, email:

engelb@ecn.purdue.edu.

Peter Vaughan, Jirka Simunek, Donald Suarez, and Dennis Corwin

Mapping CO_2 Surface Flux in an Irrigated Agricultural Area

Abstract

The 1-D Unsatchem program which calculates variably-saturated water flow, multicomponent chemical transport, CO_2 transport and heat transport was coupled to a GIS (Geographic Information System) so that calculations of water flow, CO_2 transport, and chemical transport could be conducted at an arbitrary number of point locations within a geographic area. A GIS database containing both input data and computed results was constructed for an agricultural area in the San Joaquin Valle y of California. For the growing season of 1991, a mean CO_2 surface flux of 0.0033 $m^3m^{-2}d^{-1}$ was computed at 74 locations within the field area assuming an optimal production rate of 0.0042 m^3m^{-2} d^{-1}. CO_2 flux is nearly constant for locations in fallow fields where the entire soil column remained below saturation for the duration of the simulation. In irrigated fields, CO_2 flux dropped by about a factor of f ive during irrigations because the rapid air-phase diffusion of CO_2 upwards in the soil profile was blocked due to saturation of the soil near the surface. This effect also caused higher concentrations of CO_2 below the advancing wat er front which reduced CO_2 production by suppressing respiration of soil biota. This resulted in a negative correlation between amount of irrigation water applied and the time-averaged daily surface CO_2 flux. The two soils in the st udy area, with material properties assumed to be those of a sandy-clay and a clay-loam, are not sufficiently different to affect the CO_2 flux significantly.

Introduction

Carbon dioxide surface flux is of interest at scales ranging from localized study of how particular plants process CO_2 to global studies of the carbon cycle. On a local scale, soil surface CO_2 flux was the subject of several fiel d studies in which experimental measurements of CO_2 flux were made at point locations using chambers installed in the soil (Kim and Verma, 1992; Norman et al., 1992, Kucera and Kirkham, 1971). Micrometeorological methods, in which a CO_2 sensor is mounted above ground, provide an alternative measurement technique (Verma et al., 1989). The point location studies recognized the dominant process of CO_2 production in soils as respiration of soil microorganisms and plant roots . These field studies also showed the significant variability in CO_2 flux within a given ecosystem. For example, in peatlands in Minnesota CO_2 flux ranged from 3 to 18 $gm^{-2}d^{-1}$ (Kim and Verma, 1992). In tropi cal rain forests, CO_2 flux is generally higher, ranging from 3 to 38 $gm^{-2}d^{-1}$ (Keller et al., 1986). In addition to spatial variability, CO_2 flux can have significant temporal variability as was observed during the FIFE (First International Field Experiment) after a rainstorm in 1989 when flux increased by a factor of nine (Norman et al., 1992).

To address the task of modeling both the spatial and temporal variability of CO_2 surface flux, we are reporting progress on deterministic modeling of CO_2 flux using a newly developed computer program that is coupled to a GIS. The new program, called Unsatchemgeo, is a modification of the Unsatchem program for calculating water flow in variably-saturated conditions, multicomponent solute transport, CO_2 transport and heat transport (Suarez and Simunek, 1992). Unsatchemg eo is tightly coupled to the ARC/INFO GIS, meaning that the program reads and writes directly into INFO tables.

CO_2 Production and Transport in Unsatchem

Unsatchem is a finite-element code for solving coupled flow and transport problems in the vadose zone. Processes that are modeled include water flow, multicomponent chemical transport, heat transport, CO_2 production and transport, and plant root water uptake (Suarez and Simunek, 1992). The model presumes that CO_2 production in the vadose zone is the sum of production due to respiration by plant roots and by soil microorganisms. Other factors involved in the CO_2 produ ction rate are temperature, hydraulic head, depth and the concentration of CO_2. These various factors are expressed as a product of reduction functions that muliplies an optimal CO_2 production rate (Suarez and Simunek, 1992). For e xample, the depth dependence can be expressed as an exponential function. The reduction function that is of greatest interest here is the CO_2 concentration dependence. This is expressed by the Michaelis-Menten equation,

$$q = \frac{q_{max}}{1 + \left[K_m / C_{O_2}\right]} \qquad [1]$$

where q_{max} is the maximum production rate, $C(O_2)$ is the concentration of oxygen, K_m is the Michaelis-Menten constant, defined as the concentration of oxygen when $q = q_{max}/2$ (Glinski and Stepniewski, 1985). D uring respiration the CO_2 is exchanged for oxygen on a 1:1 basis so the sum of CO_2 and oxygen concentration is presumed to be a constant value. Therefore, as CO_2 concentration increases the oxygen concentration decreases which suppresses the respiration process [Eq. 1]. The transport of CO_2 is calculated as the sum of four fluxes which include diffusive and convective transport in the air and water phases (Suarez and Simunek, 1992; Simunek and Suarez, 1993)

Coupling Unsatchem to a GIS

There are both 1-D and 2-D versions of the Unsatchem program. The 1-D version was chosen for coupling with a GIS for several reasons. First, the 1-D model only requires specification of boundary conditions at the surface and at the bottom of the soil column. Data for the surface boundary conditions are relatively easy to obtain from rainfall and irrigation records. Boundary conditions at the base of the soil column must be assumed, but this is the case for either model. For the 2-D model, one must specify a boundary condition along the sides of the plane. When symmetry considerations apply, as in a 2-D model of chemical transport between a bed and a furrow (Simunek and Suarez, 1994), these boundary conditions can reduce to simple no-flow by makin g an appropriate choice of modeling region. In a GIS context no such symmetry is generally present, so specification of the side boundary conditions is not possible without intricate field measurements. The 2-D model would also have much larger storage requirements.

Coupling the 1-D Unsatchem model to a GIS involved running the calculation for specific geographic locations. For each location, a set of input data for the model was stored in several INFO tables. During the calculation, data were read directly from these tables and calculated results were written to other INFO tables.

The point locations where the 1-D model was run, are grouped within areas called sectors. A sector is defined as an area in which the agricultural management operations of irrigation and cropping were uniform for the period of the simulation. Sectors occur within agricultural fields which are defined by physical boundaries. As an example, in one growing season, a field might be split in half along a north-south line with different crops growing in each sector. But, in the next season, the field cou ld be split East-West. Thus, for a simulation running over both seasons, the field would be split into four sectors. A spatial overlay of crop and irrigation maps can be used to generate the sector map for a given simulation. The advantage of the secto r-based approach is minimization of data storage requirements. Since all point locations within a sector have the same crop history, the crop schedule for that sector can be stored as single set of records. For the field area discussed in this paper, ir rigation of individual fields is assumed to be uniform. Thus, there is no subdivision of fields based on irrigation management practices.

The entire data set for a field area is hierarchical with attributes that apply to the entire field area appearing as single data values in tables at the top of the hierarchy. At the bottom of the hierarchy are the output data for individual nodes in the finite-element model of a soil column. For example, each node has a set of data values corresponding to each printout time during a simulation. A complete description of the structure of the database is given in Vaughan et al. (1996).

Unsatchemgeo computes the CO_2 flux at the surface for each time step at all point locations. To obtain a map of the CO_2 flux, these values must be spatially-interpolated. In this paper, spatial interpolation was carried out by o rdinary point kriging which generates a set of interpolated values at intersections on a square grid. These data sets are then processed to make maps using the raster-based GRID module of ARC/INFO.

Field Data

Irrigation

The study area consists of 37 fields in the Broadview Water District (BWD) located in the San Joaquin Valley of California. Each field is irrigated from a single source located in the southwest corner. During the growing season of 1991 nine fields rem ained fallow. The total amount of water applied to cropped fields varied from 4 to 55 cm (Figure 1). Water was applied in up to six irrigation events, each one lasting four to ten days. There is no information on precisely where irrigation water was ap plied within a field, but both the timing and the daily amount of irrigation water were carefully monitored. The actual irrigation methods included flooding of furrows, as the most common method, and sprinklers. The Unsatchemgeo model assumes that irrig ation water is applied as a uniform layer of water over the entire field.

The boundary condition of prescribed water flux, at the top of the soil column, was assumed to be the total depth of irrigation water applied during an irrigation event divided by the duration of the event. Evaporation of water from the surface was as sumed to be zero. A prescribed flux boundary condition for water flow can result in prescribed infiltration rates that are greater than may be accommodated, given a particular soil's hydraulic properties and existing water content. In this case, the mod el permits ponding at the soil surface with a resulting automatic change to a constant head boundary condition.

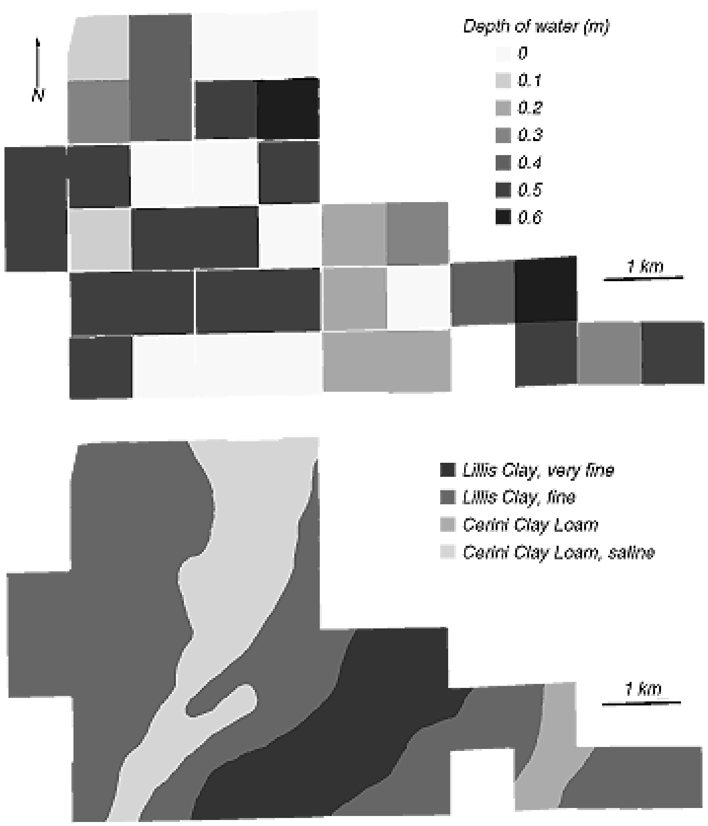


Figure 1. Broadview Water District study area. Top: Cumulative amount of irrigation water applied during the growing season of 1991. Bottom: Soil texture map based on a preliminary map prepared by the Soil Conservation Service from field data collected in 1990 and 1991.

Soil Texture Map

In 1990 and 1991, the Soil Conservation Service mapped the quadrants in which the BWD is located. There are four soil textures mapped within the study area, two are clays and two are clay-loams (Figure 1). Water retention and hydraulic conductivity da ta are not available for soils in the BWD. Instead, values of the various parameters, shown in Table 1, were taken from a short database containing these data for various soil textures (Carsel and Parrish, 1988). However, the saturated water contents sh own in Table 1 are measured values.

	$theta_s$	$theta_r$	alpha	n	K_s	rho
sandy clay	. 45	.10	.027	1.23	2.88	1.3
clay loam	.41	.09	.059	1.31	6.24	1.3

Table 1. Material Properties Relevant to Water Flow. $Theta_s$ is the volumetric water content at saturation, $theta_r$ is residual water content, alpha and n are the van Genuchten parameters, K_s is the saturated hydr aulic conductivity, and rho is the soil bulk density.

Table 2 lists other physical data values that are relevant to the calculation including the soil physical properties, the diffusion coefficients for CO_2 in air and water ($D(CO_2,air)$, $D(CO_2,water)$) and the optimal produc tion rate for CO_2 ($gamma_{s0}$).

$D(CO_2,air)$	13737.6 $cm^2 day^{-1}$
$D(CO_2,water)$	1.529 $cm^2 day^{-1}$
$gamma_{s0}$	0.42 $cm^3 cm^{-2} day^{-1}$

Table 2. Diffusion coefficients for CO_2 in air and water, CO_2 optimal production rate.

The production of CO_2 is assumed to occur by respiration of soil microorganisms exclusively. Growth of plants was not considered in the calculation so respiration of plant roots is not included. The temperature is assumed to have a constant v alue of 25 degrees C throughout the profile. These assumptions are made because the intent of this calculation is to study how irrigations and soil texture influence CO_2 flux.

Initial Conditions

In the spring of 1991, 37 fields in the BWD (Figure 1) were sampled at eight point locations within each field by the staff of the US Salinity Laboratory. Soil samples were taken at 0.3 m increments down to a total depth of 1.2 m using an auger. The w ater content of these samples was determined by gravimetric method. A separate sampling of 93 locations was done in 1993 to determine the bulk density of the soil. These locations included all but 8 of the 74 locations in the original sampling. The rem aining eight locations were in fields that could not be entered due to ongoing agricultural operations. The bulk densities for these locations were estimated by kriging. Volumetric water content, as required by the model, was calculated from the bulk de nsity and gravimetric water content. The initial concentration of CO_2 is assumed to be equal to atmospheric concentration (0.00033 cm^3cm^{-3}) down to 1.07 m. Below that depth, CO_2 concentration is assumed to inc rease linearly to 0.2 cm^3cm^{-3} at the bottom of the column (1.23 m depth).

Results of a Model Calculation

The calculation was run at 74 locations in the BWD. These locations were selected from an initial set of 315 locations because a complete chemical analysis had been conducted for soil water from samples taken at each of these sites.

Irrigation Effect on CO_2 Concentration Profiles

Figure 2 compares CO_2 concentration profiles in an irrigated field with CO_2 concentrations in a fallow field. The opacity of the curve indicates the time with darker curves corresponding to later times. There is a seven-day inte rval between curves. The first time that is represented is a week before the irrigations started. At this time the calculation had already simulated approximately 30 days of drying out. In Figure 2A the concentrations are almost constant in time and re present the development of a steady-state flow of CO_2 upwards in the profile. This steady-state condition occurs because significant air-phase porosity throughout the profile permits unrestricted diffusion of CO_2 to the surface. Th e no-flow boundary condition for CO_2 in the air phase at the bottom of the profile, coupled with an exponentially-decreasing CO_2 production rate with depth, means that CO_2 concentration is nearly independent of depth for d epths greater than about 0.6 m.

Figure 2. CO_2 Concentration profiles, later times are represented by progressively darker curves. A: Fallow. B: Irrigated.

By contrast, the CO_2 concentration profiles in the irrigated field show significant variations over time (Figure 2B). The exact sequence of curves can be determined by looking at CO_2 concentrations at a depth of one meter. At th is depth, increasing CO_2 concentration has a 1:1 correspondence with the times represented by the curves. The first profile is essentially the same profile as occurs in Figure 2A, but the second profile has a peak at about 0.06 m. This peak i s caused by a restriction on the upwards flow of CO_2 due to the displacement of air-phase porosity by water while an irrigation is occurring. Thus, the sequence of irrigations and drying episodes leads to oscillations of CO_2 concent rations as shown by Suarez and Simunek (1993). In this model, there is an upper limit on the concentration of CO_2 imposed by Michaelis-Menten equation, so CO_2 concentrations are not likely to be much higher than the maximum shown in Figure 2B.

Effect of Soil Texture on CO_2 Concentration

Two locations, in an irrigated field, occur in soil units that were mapped as clay and as clay-loam. By assuming an identical initial water content profile for both locations, the effect of the difference in soil texture is seen to be negligible for b oth water content and CO_2 profiles (Figure 3). Larger quantities of

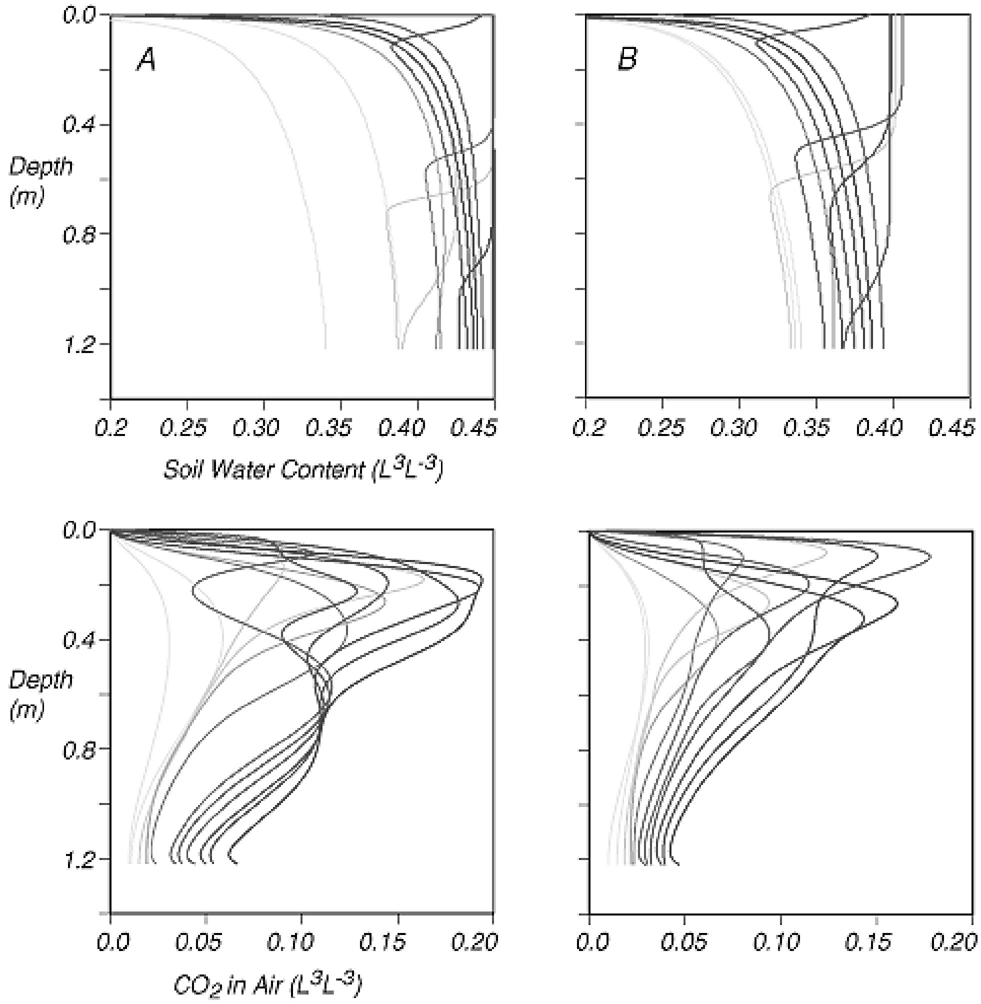


Figure 3. A: Water content (top) and CO_2 concentration (bottom) vs. depth for sandy-clay. B: Water content (top) and CO_2 concentration (bottom) for clay-loam.

irrigation water were applied to this field than most others in the study area and the calculation indicates that at least a portion of the soil profile was saturated for about 30% of the growing season. High concentrations of CO_2 (> 0.15), developed during this period. These high concentrations persisted for at least 3 weeks after drying reduced the water content of the entire column below saturated values. A similar computed buildup of CO_2 occurred at locations in irrigated f ields within all four soil units.

Flux of CO_2

CO_2 flux is essentially constant in a fallow field because there is a steady production rate and unimpeded diffusion of CO_2 to the surface through the air-filled pores. In irrigated fields, the

CO_2 flux drops by a fact or of five during a flood irrigation (Figure 4). Immediately after an irrigation, the CO_2 flux goes through a transient period, in which the flux rises to a maximum value, and then goes through a minimum before leveling off. The maximum is li kely caused by the accumulation of CO_2 in soil layers below the saturated zone and its subsequent release after an irrigation had terminated. The reduction of CO_2 flux due to irrigations translates to a reduction in CO_2 c umulative flux during a growing season. Ultimately, this reduction in the cumulative CO_2 flux results because high CO_2 concentration suppresses the CO_2 production as indicated by equation 1.

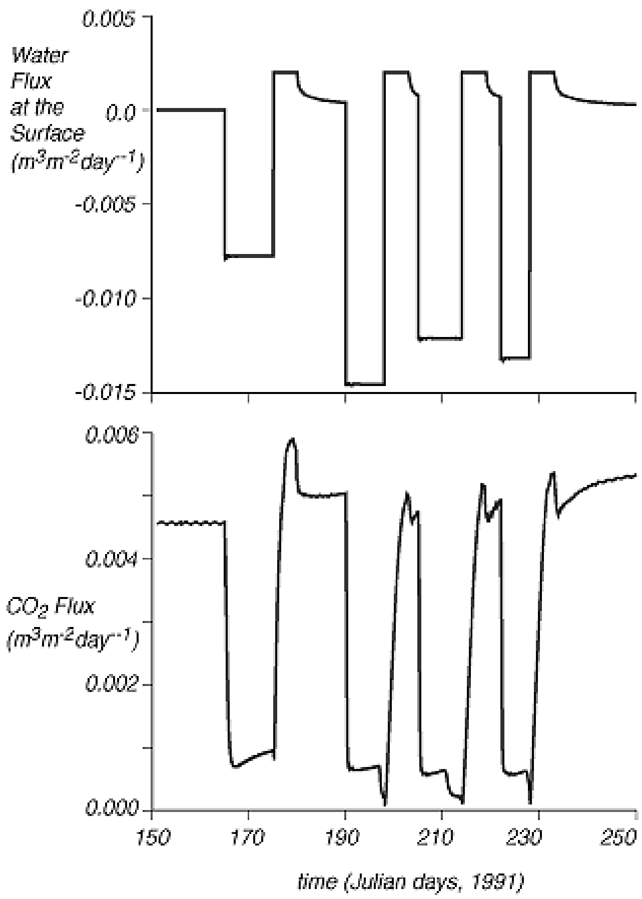


Figure 4. Plots of surface water flux and CO_2 flux over the same time period.

A map of CO_2 surface flux was obtained by kriging the cumulative flux at 74 locations within the BWD (Figure 5). The mean daily CO_2 surface flux for the 74 locations is 0.33 $cm^3cm^{-2}d^{-1}$ and the vari ation is from 0.05 to 0.48 $cm^3cm^{-2}d^{-1}$.

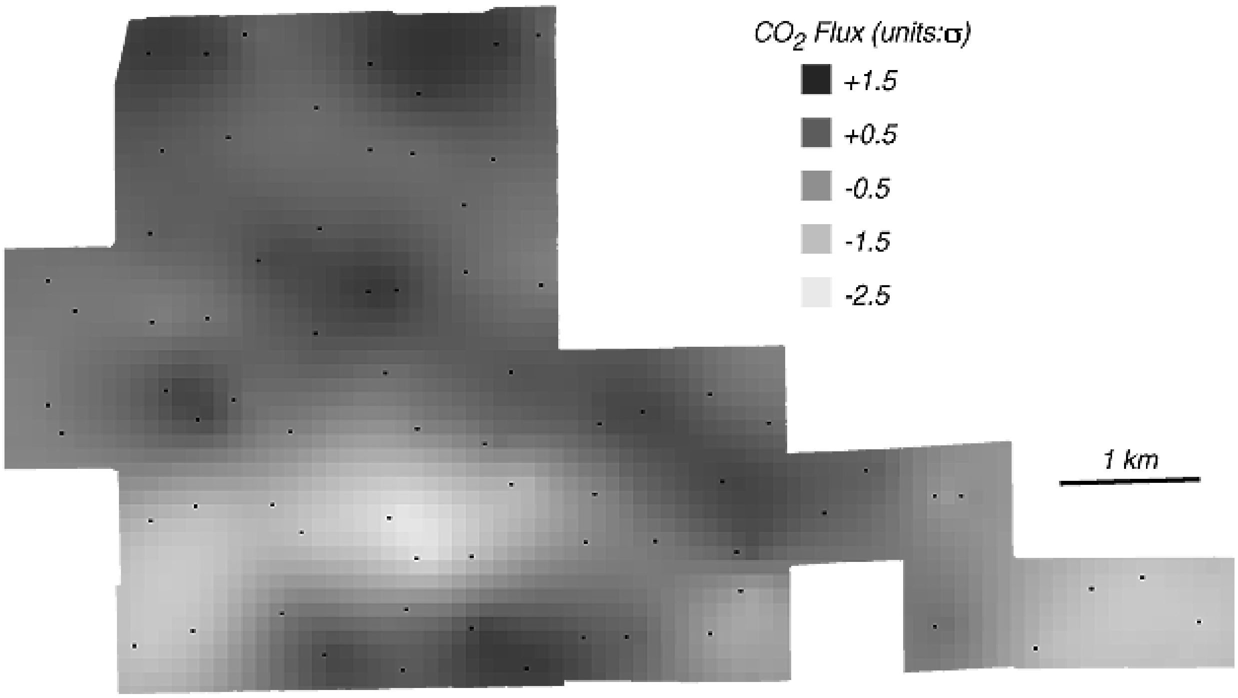


Figure 5. The raw data for this plot are the time-averaged daily surface CO_2 fluxes during 1991. These data were normalized to zero mean and unit variance to make this map which shows variation in units of standard deviation. The s mall squares designate point locations where the calculation was run.

The spatial variation of surface CO_2 flux has a negative correlation with amount of irrigation water applied as can be seen by comparing Figures 1 and 5. For example, the fields having the smallest amounts of irrigation water applied are th e areas of highest CO_2 flux. The two fields receiving the largest amounts of irrigation water, however, are not in areas where CO_2 flux is lowest. Another factor that is important in controlling surface CO_2 flux is the in itial water content. The two fields in the southeast corner of the study area had initial water contents near saturation. Even though the irrigation amounts applied to these fields fall in the middle of the range, the CO_2 flux is low because soil in these fields was at, or near, saturation for much of the period of the simulation. The other factor, investigated in this study is soil texture. By comparing Figures 1 and 5, there is no indication of spatial correlation between time-averaged CO_2 flux and soil texture.

Conclusion

The Unsatchemgeo program can be applied to modeling CO_2 production and transport in the vadose zone within a spatial context utilizing a GIS. Simulation of water flow, CO_2 production and CO_2 transport for an irrigated, agricultural area in the San Joaquin Valley suggests that significant variations in surface CO_2 flux may be occurring due primarily to variations in amount of irrigation water applied. During an irrigation event, CO_2 transport to t he surface is reduced by approximately a factor of five because saturation of the soil, near the surface, blocks diffusive transport of CO_2 in the air phase causing concentrations of CO_2 to increase at greater depths. This has the e ffect of suppressing respiration by soil biota which causes lower overall CO_2 production rates. Thus, the time-averaged CO_2 flux is lower in areas where irrigations cause saturation of the near-surface soil.

One potential application of Unsatchemgeo is the identification of areas where field measurements of surface CO_2 flux would be most likely to show significant variations over relatively short distances. Such an area in the BWD would consis t of several fields near the southern boundary (Figure 5). Field measurements are needed for comparison with the computed results presented here before a firm conclusion can be reached regarding acceptability of the model for CO_2 flux predicti ons in the BWD.

References

Carsel, R.F., and R.S. Parrish (1988) Developing joint probability distributions of soil water retention characteristics. *Water Resour. Res.*, 24:755-769.

Glinski, J., and W. Stepniewski (1985) Soil Aeration and Its Role for Plants, CRC Press, Boca Raton, Florida.

Keller, M., Kaplan, W.A. and S.C. Wofsy (1986) Emission of N_2O, CH_4 and CO_2 from tropical forest soils. *J. Geophys. Res.* 91: 11791-11802.

Kim, J. and S.B. Verma (1992) Soil surface CO_2 flux in a Minnesota peatland. *Biogeochemistry*, 18:37-51.

Kucera, C.L. and D.R. Kirkham (1971) Soil respiration studies in tallgrass prairie in Missouri. *Ecology*, 52:912-915.

Norman, J.M., Garcia, R. and S.B. Verma (1992) Soil surface CO_2 fluxes and the carbon budget of a grassland. *J. Geophys. Res.* , 97:18845-18853.

Simunek, J. and D.L. Suarez (1993) Modeling of carbon dioxide transport and production in soil: 1. Model development. *Water Resour. Res.* , 29:487-497.

Simunek, J. and D.L. Suarez (1994) 2-Dimensional transport model for variably saturated porous media with major ion chemistry. *Water Resour. Res.*, 30:1115-1133.

Suarez, D.L. and J. Simunek (1992) UNSATCHEM code for simulating one-dimensional variably-saturated water flow, heat transport, carbon dioxide production and transport, and multicomponent solute transport with major ion equilibrium and kinetic chemistr y. U.S. Salinity Laboratory Technical Report #129 (available from U.S. Salinity Laboratory).

Suarez, D.L. and J. Simunek (1993) Modeling of carbon dioxide transport and production in soil: 2. Parameter selection, sensitivity analysis and comparison of model predictions to field data, *Water Resour. Res.* , 29:499-513.

Vaughan, P.J., Simunek, J., Suarez, D.L., Corwin, D.L. and J.D. Rhoades (1996) Unsatchemgeo: Modeling water flow and multicomponent transport in a GIS context. *Soil Sci. Soc. of Am. J.* (Special Publication for Boyoucos Conference: *Application s of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone*), in press.

Verma, S.B., Kim, J. and R.J. Clement (1989) Carbon dioxide, water vapor and sensible heat fluxes over a tallgrass prairie, *Boundary-Layer Meteorol., 46:53-67*.

Author Information

Peter Vaughan, Donald Suarez, Jirka Simunek, Dennis Corwin
US Salinity Laboratory, ARS-USDA
450 W. Big Springs Road
Riverside, CA 92507
email: {pvaughan, dsuarez, jsimunek, dcorwin}@ussl.ars.usda.gov

Model Comparison of Solute Transport Models at Regional Scale

Alejandro Mateos, Palmaven S.A., Stephen D. DeGloria and R. Jeff Wagenet

Abstract:

Deterministic simulation models are being developed and used for understanding and quantifying major transport processes involved in the leaching of solutes from soils under various climate conditions, soil properties, and agricultural practices. Such models provide useful tools for water management and chemical application studies of non-point source pollution at regional scales. In contrast, soil information at regional scales is uncertain, in the sense that it can have many possible values.

When uncertain input data is used in simulation models, the model output data can also be expected to be uncertain. If a simulation model is used to support land and water management decisions, the uncertainty of model predictions can be of particular importance in defining the level of confidence of using the model output to influence the decision making process.

In this study of a systematic approach to assessing the impact of soil data uncertainty of leaching model predictions was developed. The approach based on Monte Carlo simulations was used to estimate the probability distribution of leaching model outcomes resulting from the uncertainty of soil survey information. Map unit variability was characterized by the data uncertainty of individual soil components. The proposed method was found easy to conceptualize and to implement.

Additionally, two simulation approaches to water flow and solute movement using deterministic simulation models that differ in algorithm complexity, one being a mechanistic-type and the other a functional-type, were compared using the uncertainty analysis approach developed here. The results of the comparison suggested that the simpler functional modeling approach of water and solute movement has promise as an alternative approach with comparable precision to more complex mechanistic models when used for regional scale studies.

Current applications were reviewed and recommendations made for the appropriate use of and improvements to soil survey data for modeling pesticide leaching at regional scales.

Cornell University

Individual-based Modeling SESSION and SWARM Workshop

Chris Langton

Abstract: **Modelling Complex Adaptive Systems**
Systems composed of many interacting "agents" can exhibit quite complex behaviors, especially if the agents are "adapting" through time by altering the rules upon which they base their actions. I will discuss problems in understanding such systems and some techniques for addressing them.

Chris Langton
Research Scientist
Santa Fe Institute (SFI)
1399 Hyde Park Road
Santa Fe, NM 87501
phone: 505-984-8800
email: cgl@santafe.edu

Michael F. Goodchild

Directions in GIS

Introduction

The purpose of this paper is to discuss research directions in GIS. In doing so I will deliberately broaden the meaning of "GIS" to include a wide range of activities within the broad rubric of digital geographic information, since it seems to me that "GIS" is now generally used in that broad sense, rather than in the narrower sense of a software system designed specifically to store, retrieve, and analyze existing geographic information (Maguire 1991). We seem to be reaching a point where digital technology is encountered in almost all aspects of the communication of information, and the same is true of geographic information. It no longer seems important to ask whether information is digital or not--rather, the important questions seem to concern the degree to which the digital format imposes itself on the information, forcing us to modify, transform, or otherwise alter the information in order that it can be handled and communicated in digital systems.

GIS research is now very broad, ranging from investigations into how people think and reason with geographic information, so that better systems can be designed that are easier to use, to studies of the legal and intellectual property issues raised by widespread sharing of geographic information. The work of the U.S. National Center for Geographic Information and Analysis has encompassed much of this range (for more information on the work of NCGIA see http://www.ncgia.ucsb.edu), as have recent conferences on GIS research such as the International Symposia in Spatial Data Handling (Waugh and Healey 1994), the international Symposia on Spatial Databases (Egenhofer and Herring 1995), and the Conferences on Spatial Information Theory (Frank and Kuhn 1995). In this paper I can hope to cover only a small part of that range, and the subjects I discuss in what follows reflect certain obvious personal biases, toward NCGIA, the University of California, Santa Barbara, the Alexandria Digital Library project (http://alexandria.sdc.ucsb.edu), and more generally the U.S. University Consortium for Geographic Information Science (http://www.ucgis.org), and the European Science Foundation's GISDATA program (http://www.shef.ac.uk/uni/academic/D-H/gis/gisdata.html).

The paper begins with the concept of a life cycle for geographic data, and discusses the changes that are occurring at various stages along it. The data life cycle is a convenient organizing mechanism for current GIS, given the recent trend to extend digital geographic data handling to all aspects of the cycle from initial observation through to eventual archiving. I then provide a description of the Alexandria Digital Library, an effort to provide the services of a map and imagery library over the Internet, and to exploit the power of geographic location as a means for organizing information. Subsequent sections explore some of the general issues raised by Alexandria, in the areas of metadata, information granularity, and scale. The transition to a digital world is causing us to question many of the traditional ways of doing things, and to ask what aspects of this legacy should be preserved, and what abandoned. The paper then moves to a discussion of the major impediments to GIS, and directions in which GIS might evolve to avoid them, specifically in the context of data models, which I argue lie at the root of GIS capabilities. This leads to a discussion of interoperability in GIS, which is one of the keys to improvement in the ease of use of the technology.

The data life cycle

A traditional view of GIS, reflected in much early writing in the textbooks of the field, is that its purpose lies in building a digital representation of some existing set of geographic data, thus allowing the data to be subjected to analysis. Spatial analysis is often seen as the primary purpose of GIS, and the results of analysis as being used as the basis for decisions (Cowen 1988). In this view, the key elements of GIS include the ability to convert to digital form, store, manipulate in a fashion analogous to the calculator, and report. Much has been written about the kinds of functions needed to support these activities (Maguire and Dangermond 1991), and on related issues such as the accuracy of each stage. Comparisons have been drawn between these functions of spatial analysis and those of a statistical package--it has been said that GIS is to spatial analysis as a statistical package is to statistical analysis (Goodchild 1987). Underlying the entire framework is the notion that the value of computers lies in their ability to process numeric, and sometimes alphanumeric, information much more rapidly than can a human being. Indeed, the initial Canada Geographic Information System, one of the earliest GIS, had no capabilities at all for visual display, since its intended purpose of numerical analysis of maps could lead only to tabular, numeric output (Tomlinson 1967).

In this view, the source of information for GIS is commonly the map, and the conversion of its contents from paper to digital form through digitizing and scanning is a key GIS operation and one that is ultimately very costly. But once in digital form, the high speed with which the data can be processed normally provides the benefits needed to justify the conversion. However, the few minutes needed to process the data compare poorly to the many months that are often required to prepare the data for processing. Ironically, therefore, GIS-based projects when considered in their entirety are often significantly slower than ones completed using more traditional methods.

Several factors have caused this view of GIS to change in the past five years, and have led to its broadening to include many more data-related activities. First, GIS is no longer an activity confined to the desktop. It is now possible to obtain sophisticated processing power and storage in a portable device weighing less than a kilogram, and to think of GIS as an activity that can take place in a car, or almost anywhere. The concept of "field GIS" is now exemplified by many new applications that have emerged in the past five years as digital technology has become smaller and more compact. GIS is now the basis for the new technology of precision agriculture, a term used to describe the use of geographic information technologies to improve greatly the precision with which land is farmed. Rather than make decisions about entire fields, farmers can now determine appropriate levels of seeding, fertilizer, and pesticide applications based on detailed information collected from continuous monitoring of crop yield by harvesters, air photographs, and detailed soil surveys. Precision agriculture is favorable to the environment, because levels of application of pesticides and fertilizers can often be reduced based on detailed knowledge of soil conditions; and it is also economic, since the improved yields and cost savings are both reflected in improved income to the farmer.

Field GIS is also emerging in the form of new software designed to support collection of data rather than its analysis. The "field notes" form of GIS often includes a "backcloth", typically a georeferenced image, on which additional information can be located by the field scientist. The software is installed in a portable computer, often with an associated GPS receiver for accurate geopositioning, and maps and sketches created by the observer can be uploaded to a conventional desktop GIS either by wire or by radio communication. Such systems are now commonly used to collect information from forestry transects, or in routine maintenance of utility networks. Another form of field GIS, installed in a van and coupled to kinematic GPS, can be used to produce accurate surveys of road networks and to capture associated digital pictures of the state of road pavement or signage (Novak and Bossler 1995). In all of these examples field GIS is allowing us to see GIS more as a tool for data collection and the construction of geographic databases than for their analysis--the database becomes the product of GIS, rather than its input.

Second, GIS is rapidly becoming a technology that impacts not only our professional lives, but our everyday ones as well. Rather than being confined to sophisticated tasks of inventory and analysis, GIS is now appearing in the form of simple mapping capabilities attached to spreadsheets, as exemplified by the recent extensions of Microsoft Excel into simple cartography. Advertisements for simple GIS products appear in airline magazines, and GPS units are available in electronics stores. Car rental agencies now offer in-car navigation systems. In the next few years, we can expect many other applications of GIS for the mass market to appear, including systems that automatically map the locations of accidents based on explosions of car air bags, and trigger appropriate emergency response.

This interest in GIS is being fuelled in part by the availability of low-cost hardware, software, and data; and in part by a growing level of interest in geography, geographic information, and the kinds of thinking and reasoning processes associated with space. It is said that a picture is worth a thousand words; and that maps store and communicate information much more efficiently than text; and that the human eye and brain are marvellously adapted to rapid analysis of visual presentations.

A third point concerns the role of digital geographic technologies in communication. In the traditional data life cycle, technology can be seen as a means of transmitting the knowledge gathered in the field to its eventual users. Decisions must often be made based on knowledge gathered in the field, in circumstances where field scientist and decision-maker are different people, separated by great distances and also perhaps by differences of discipline and experience. In such circumstances the map or spatial database becomes the communication channel, and its contents the means by which knowledge is transmitted. The paper map is a very effective means of communicating certain types of information, but it can severely restrict the communication of other types. For example, paper maps require us to flatten the world; they confine us to a fixed, uniform resolution; and they tend to force information that is frequently vague or fuzzy into precise categories. Such restrictions are a major constraint on the traditional data life cycle, since maps have been the primary repository and communication medium for geographic information. In the digital era, there is the potential to remove many of these constraints, and thus to widen the communication.

The fourth factor affecting the data life cycle concerns data access, and the changes occurring in the architectures of computing. Unlike the early days, when computing was limited to mainframes, and later developments of interactive access via "dumb" terminals, today's computing environments include local and wide area networks (LANs and WANs), the Internet, high-bandwidth communication channels, clients and servers. Such new architectures allow radical restructuring of the various stages and roles in the data life cycle. With client-server technology, it is possible for the originator of the data to become its permanent custodian, thus ensuring that it is as current as possible, and that no problems arise through the existence of different versions in different locations. Data can also be documented and described at its source, and the documentation updated and maintained digitally through each stage in the data's transformation.

With current object-oriented technology, it is possible for important processes to be encapsulated with data. For example, it may be useful to encapsulate suitable display software, so the potential user can see the data in spatial form. Much data gathered at sample points, such as weather stations, will frequently need to be interpolated for analysis, and encapsulation allows its originator to make decisions about how best to interpolate. Finally, processes must be encapsulated with data if the goals of interoperability are to be achieved, since data must be capable of describing itself, and how it should be processed, to a variety of host systems.

In summary, there are many reasons to believe that geographic information technologies will play a role at every step in the life cycle of geographic data, not just in its analysis. The data life cycle concept allows us to see the entire process of data handling, from creation to archiving; to ask questions about the efficiency of communication of knowledge; to look at the roles of various participants; and to ensure that information is passed as effectively as possible between people and techologies that traditionally may have found it very difficult to share and communicate because of differences in terminology, formats, and disciplines.

The Alexandria Digital Library

Many of these issues can be illustrated within the framework of the Alexandria Digital Library project (ADL). Named for the classical library of Alexandria, ADL aims to build a distributed digital library for spatially referenced information, including services for constructing the library's collection (ingest), cataloging, retrieval, and use. All of these functions are to be distributed, and accessible via the Internet. We define spatially referenced information as any information referenced to a two-dimensional frame, of which geographical reference to the surface of the Earth is ADL's primary example. While one can think of ADL as an effort to build a map and imagery library on the Internet, its objectives go further in seeking to exploit geographic location as a method for indexing and retrieving information; and to make maps and images part of the mainstream of the digital library of the future.

Although the idea of a digital library may seem far-fetched, the technology already exists to store the entire collection of humanity's published text in digital form--and almost all text now published is available in digital form. On the other hand maps and images are comparatively voluminous, and a single day's output of the EOS generation of satellites may exceed the information content of many of today's research libraries. Nevertheless, the advantages of universal access to library services via the Internet are very attractive.

ADL is one of six digital library projects being funded by the U.S. National Science Foundation (NSF), the Advanced Research Projects Agency (ARPA), and the National Aeronautics and Space Administration (NASA). The concept of ADL was developed at the Map and Imagery Laboratory at the University of California, Santa Barbara (UCSB), over several years, leading to major funding beginning in October 1994. ADL has many participants and partners, including the Departments of Geography, Computer Science, and Electrical and Computer Engineering at UCSB; the Library of Congress and the U.S. Geological Survey; Digital Equipment Corporation; Environmental Systems Research Institute (ESRI); and many others. Projects like ADL provide exciting opportunities for the GIS community, because they offer links to the vast resources and experience of libraries and library science.

During the first six months of funding, ADL developed its Rapid Prototype (RP) based on existing technology, and designed to act as a straw-person for further discussion. The RP is a standalone system, based on ESRI's ArcView, the interface language Tcl/Tk, and the database management system Sybase. The RP was completed in March, 1995, and has been distributed on CD for evaluation. Following the RP, ADL moved to the development of a universally accessible server implementation based on the Internet and the World Wide Web. A Web prototype was developed by November 1995, and is currently in beta release (for details, see http://alexandria.sdc.ucsb.edu). The number of digital data sets now accessible via the Web prototype is on the order of 105.

The strategy used in the design of the Web prototype is based on some simple assumptions about the nature of searches for spatially referenced information. Almost all such searches start by specifying location; subsequently, the search is refined to certain subjects or themes, ranges of dates, data formats, and levels of geographic detail. Thus a user might request information on Taiwan, covering surficial geology, as recent as possible, in ARC/INFO format, and at a scale of 1:100,000 or better. To accommodate this search strategy, the user of the Web prototype first encounters a world map (other base maps can be used as appropriate). Pan and zoom features are available to focus attention on any area, and additional features and detail become visible on the world map as the user zooms. A query area can be defined, and used to search the ADL catalog for suitable data sets. Other features allow the query to be restricted to certain dates, themes, data formats, etc.

ADL has adopted a standard format for its catalog that is as compatible as possible with both library and GIS communities. On the library side, we have made the format compatible with U.S. MARC, a standard for digital catalogs. On the GIS side, ADL has adopted a subset of the Federal Geographic Data Committee's (FGDC) Content Standard for Geospatial Metadata (http://geochange.er.usgs.gov/pub/tools/metadata/standard/metadata.html), which specifies some two hundred possible descriptive fields. The idea of using a "core" set of metadata elements in the digital library context rather than an exhaustive description is now widely recognized (see, for example, http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html).

Although both the RP and the Web prototype allow the user to specify the location of a query, typically as a "bounding box" consisting of two latitudes and two longitudes, most users of map libraries specify their needs not in coordinates but through place-names. The great flexibility of place-names, and the hierarchical structure they imply, have made them very difficult to use as bases for library catalogs, and this is one reason why maps and images have never been part of the library mainstream. For example, it is possible that a query about Santa Barbara might be satisfied by information cataloged as "Southern California", or as "Goleta", a suburb of Santa Barbara. These horizontal and vertical linkages make it very difficult to access information by place-name.

In the ADL Web prototype, we have implemented a large gazetteer, of about 5 million entries (a gazetteer is defined as an index linking place-names to geographic locations, with place-names as the primary key; most atlases include gazetteers linking place-names to pages in the atlas). This allows the user to search by place-name, to select among various places of the same or similar names, and to see the basemap positioned and scaled appropriately.

The use of the gazetteer gives the Web prototype several powerful features. Besides the ability to specify the location of search by name, the user can also achieve a certain level of "content-based search", or search of the actual contents of the data set rather than the contents of its documentation or metadata. Thus it is possible to search for information on Santa Barbara even though the words "Santa Barbara" do not appear in the data set's catalog entry--the simple geographic intersection between the location of Santa Barbara, as determined from the gazetteer, and the data set's bounding box, are sufficient. The gazetteer can also be used to broaden or narrow a search. For example, by determining from the gazetteer that Santa Barbara is within the State of California, one might broaden a search for information on Santa Barbara to data sets covering the entire state--and similarly search can be extended to Santa Barbara's geographic neighbors if these can be determined from the gazetteer.

There are many problems associated with implementing a gazetteer. First, and perhaps most problematic, is the lack of information on feature extents, since most gazetteers provide only a point location for a feature. This is acceptable for a small, point-like village, but is inadequate for extended features such as long rivers or large political entities. Standard protocols for such features, such as the use of the river mouth as representative point, are not sufficient for the purposes of ADL. It would be impractical to add geometric extents to every feature in a large gazetteer. Instead, we are using various rules, in conjunction with available information such as feature type, to try to infer feature extent in as many cases as possible, and in other cases inviting the user to make appropriate decisions and choices.

While we are far short of having an operating digital library, projects such as ADL allow us to see aspects of what a digital library might offer, and to anticipate some of its problems. Many impediments now stand in the way of offering the full services of a library over the Internet, and one that provides access to all types of information irrespective of the somewhat restrictive legacy of the traditional library. In the next sections I discuss three of them: metadata, the issue of data granularity, and the specification of level of geographic detail.

Metadata

Metadata is normally defined as data about data--the information that allows us to find, handle, browse, read, and understand the contents of a data set. The concept of metadata embeds a range of metaphors, from the library card catalog, through the written data documentation needed by early generations of computer technology, to the handling instructions that appear on the outside of a mailed package. All three are valid, since metadata must perform elements of all three functions. Metadata must be digital, must travel with or ahead of the data, and must be preserved as the data are transformed by various kinds of processing, and if possible updated appropriately. For example, a change of projection process should cause a corresponding modification of the associated metadata.

While all three metaphors convey a sense of the functions inherited by metadata from the pre-digital era, none of them reflect a world that truly takes advantage of digital technology. One of the most significant advantages of digital technology is the opportunity to rethink the data life cycle, and to add to the value of data by making it shareable and useful to much larger populations of users. In this context metadata provides the opportunity to explain the meaning of data, allowing users who may have had nothing to do with its creation, and may not be familiar with the terminology and practices of the community that created it, to put it to use. Thus Francis Bretherton has defined metadata as "that which makes data useful".

In this sense, the value of metadata to potential users is related to the existence of a common language, and to the uses to which the data will be put. A user who needs information on the positional accuracy of a data set is not necessarily helped by knowing the serial number of the digitizer on which it was created, or the date, unless these items of information can be translated into terms understood by the user, such as the 90th percentile of positional error. Items such as the serial number or date thus reflect a producer's approach to metadata, where the main value of such information is for production control, rather than the needs of the user.

In essence, the metadata needed to support digital libraries must serve as the basis for a dialog between producer and user, in a common language understood by both. It must reflect as much the skills and experience of the user as of the producer. Since the skills of library users range widely, from "spatially aware professionals" (SAPs) to elementary school students, metadata must be able to present many different faces.

In ADL, we are experimenting with studies of the process that occurs when a user enters a map library, in order to gain understanding that can guide the design of the digital library. Ethnographers are recording the process on videotape for analysis, and are finding that ADL is already creating its own culture, with a language that is distinct both from that of the traditional map library and that of the computerized library card catalog.

We have concluded that the only feasible approach to metadata in this context must be hierarchical. SAPs will want to search the library using very cryptic languages that convey very dense meaning; elementary school children use much less compact languages. Correspondingly, ADL must support a hierarchy that ranges from the extreme detail of the FGDC metadata standard to a small number of "core" metadata elements likely to be understood by all. This is a very different approach to those of card catalogs, data documentation, and single-tier metadata standards, but it is necessary if the library is to be as useful as possible to as many as possible.

Granularity

In the traditional library, with its stacks of books, information is organized around the bound volume as the fundamental unit. Very few elements of traditional library catalogs address either groups of books, or portions of a book's contents; the discrete enumeration and cataloging of volumes takes advantage of the physical nature of the book as a unit. Only in the case of serials is there any significant extension to a more general concept of information granularity. Moreover, this same principle can also be applied to a map library's map sheets, standard-sized images, and atlases.

In the digital world, these concepts of granularity are no longer as easily followed. The contents of a single topographic map sheet are likely to be stored in several files, each corresponding to a layer; the full representation of a standard 1:24,000 U.S. Geological Survey topographic map may run to tens of files in certain GIS representations, and there is little reason to store them in one aggregate. Equally, the digital world allows us to escape from the constraints of map sheets, and to explore the feasibility of a continuous, "seamless" view of the world. Why should there be 50,000 individually cataloged 1:24,000 quadrangles, rather than one seamless data set, and why should Landsat imagery be stored in awkwardly shaped "scenes"? Other problems of granularity arise in the case of maps embedded in atlases and books, or geographic data sets in CD collections.

In short, the 1:1 model of the card catalog--one card per book--is overly restrictive and constraining, and not defensible in the digital era. Instead, we need to develop concepts of information abstraction and generalization that span the entire range from the individual feature, through the feature class and map sheet, to the seamless layer or mosaic. The concepts of data and metadata are no longer as separable as they were in the past.

Level of geographic detail

Of the four main dimensions of a search for spatially referenced information, it is the level of geographic detail that has caused us the greatest difficulties. Traditionally, the cartographer has defined the level of detail shown on a map largely through its metric scale: a scale of 1:24,000 defines for the U.S. Geological Survey not only the ratio of distance on the printed map to distance on the ground, but also the positional accuracy, the set of features depicted, and the level of cartographic generalization applied to them. But metric scale has no meaning for data in digital form, since unlike paper maps there are no distances to be measured in a digital computer's storage.

Although metric scale has no meaning, the other properties related to scale survive the conversion to digital form, including positional accuracy, content, and generalization. Thus metric scale has proved useful, and is often specified in metadata. We know, for example, that data labeled "1:24,000" will show most city streets and name some of the most important of them, but will not include most individual buildings. It will show many streams and creeks as single lines, and will have a positional accuracy consistent with the National Map Accuracy Standards (NMAS) for 1:24,000 data, that is, a 90th percentile of error of approximately 12m. But since the four scale-related elements are now decoupled, there is no way of knowing what it means to say that a U.S. Geological Survey Digital Orthophoto Quad (DOQ) has a "scale" of 1:12,000. Clearly this is not the metric scale. Nor is it indicative of spatial resolution, which is 1m for these data sets, or of the features contained, which are not explicitly identified in this raster data. It turns out that the reason for specifying a "scale" of 1:12,000 has to do with the positional accuracy of a DOQ, which is about 6m, and thus compatible with the NMAS for maps of that scale. But this kind of reasoning is largely arbitrary, and very confusing to inexperienced users, who might ask why positional accuracy is not simply specified in m.

For SAPs, it makes sense to specify level of geographic detail through as many of its dimensions and indicators as possible--a DOQ's metadata record should include its positional accuracy, spatial resolution, and content, as if these are largely independent properties. On the other hand inexperienced users may be happier with a looser definition that exploits a suitable metaphor. The Microsoft Encarta Atlas specifies the scale of a display by using the metaphor of a human eye positioned at some distance above the surface of the Earth--descent produces more detail, and ascent produces less. Of course this is not precise, and it says little about positional accuracy, but it may be fully adequate for the inexperienced user who has not thought much about geographic data and its fundamental concepts.

Summary

In summary, ADL provides us with a prototype or straw-person which we can use to explore some of the implications of digital library technology in the specific context of maps, images, and other spatially referenced information. One of its most powerful concepts concerns the use of geographic location as a means for organizing and retrieving information. In the traditional library geographic location has played little role in search, for several reasons. In fact, one can find information by its location only if a suitable place-name appears in one or more of the author, title, or subject of a book, and then only by some form of keyword search. This is very restrictive--it means, for example, that one cannot use the catalog to find information on Paris in a guidebook cataloged under "France". In the form of search, browse, and retrieval implemented in a digital library such as ADL, however, the geographic key becomes one of the most useful, particularly for information that has a geographic "footprint". At this point we are restricted to footprints that are both "crisp" and singly-bounded, and force them still further into the form of bounding boxes. But in future versions of ADL we expect to be able to deal effectively with "fuzzy" and multiple footprints, and to link information through a common geographic key. In a world that is increasingly in need of new ways of organizing its ever-growing information base, such extensions offer great potential.

The importance of legacies

Technologies like ADL are caught in a fundamental contradiction, between the need to create digital environments that seem familiar, by making use of familiar metaphors, on the one hand, and the need to exploit fully the power of the digital environment on the other. For example, it may be important to provide the "look and feel" of the card catalog--but this is basically incompatible with the ideas expressed in the previous section. Old ways of thinking provide comfort in the digital world, and GIS and ADL are full of them--GIS is often explained as "having a map in a computer", and digital libraries as the digital equivalent of books on shelves. In ADL and in GIS generally we face the need to find a balance, by using the legacy of previous technologies to provide familiar landmarks, while trying to escape the constraining aspects of legacies. Sometimes this dilemma is resolved by selecting one direction over the other, as when automated cartography is used to produce maps that are indistinguishable from their pre-digital parents. In other cases new technologies exploit new concepts to such a degree that they appear hostile to the user, and require lengthy training prior to use.

This argument is usually presented in the context of the transition from analog to digital, but it can apply equally to the replacement of old digital technologies by new ones. Early versions of GIS technology were very restrictive, and we will continue to find new ways of broadening the base of GIS and removing some of the earlier constraints. In this sense GIS will never be a perfect technology, and must continuously reinvent itself in order to overcome the constraints of the past. In the next four sections I present four ways in which I believe GIS can reinvent itself, and indeed must do if it is to continue to grow. Each of these four reflects a current area of GIS research, where recent results suggest the possibility of substantial progress.

One of the great advantages of GIS is its ability to store and manipulate a wide range of types of geographic data. Any such software environment that seeks to provide a range of capabilities and functions is fundamentally driven by the data model that it implements--by the set of options provided by the underlying database for the storage and retrieval of information. Thus a word processor offers a vast assortment of capabilities and functions for written text, but would be of little use for processing raster images. Similarly a spreadsheet implements the functions that operate on simple rectangular tables of information. A data model is defined as the set of entities and relationships used to build a representation of some real-world phenomena, and for GIS these are the entities and relationships needed to represent variation on the Earth's surface.

A growing literature has drawn attention to the role of the data model, and to possible ways in which GIS data models might be extended, made more versatile, or otherwise broadened to support a wider range of GIS operations. But ultimately it is the choice of data model by the GIS designer that determines what the GIS can be made to do. Goodchild (1992) has identified three broad classes of data models--those needed to represent continuous fields, discrete objects embedded in a two-dimensional space, and variation over linear networks embedded in two-dimensional space. But although these three classes capture almost all data models currently implemented, the level of user interaction is with a lower conceptual level, where geographic variation is represented in terms of geometric objects--points, polylines, polygons, and raster cells--and their attributes and relationships. No less than six distinct data models are used to implement the concept of a continuous field--a raster of cells, regularly spaced sample points, irregularly spaced sample points, digitized contours, polygons forming an irregular tesselation of the plane, and triangulated irregular networks (TINs).

Interoperability between data models

This problem of a plethora of data models addressed in the previous paragraph is one reason why current GIS has the reputation of being difficult to use. For example, a set of points stored in a GIS with associated attributes may represent very different things to the user, as representations of very different data models, and the associated meanings of different functions and processes are different as well. An irregularly spaced set of points may represent a sample of weather stations, and an attribute such as mean temperature may be used to create a continuous surface by using a process of spatial interpolation. But the same function of spatial interpolation would be meaningless if the points represented cities, and the attribute was population, or average income. Because interaction with current GIS commands is almost always with the primitive element, such as the point, rather than with what the point represents, it is necessary for the user to maintain a high level of conceptual involvement with the application. The GIS is incapable of warning the user when something meaningless is being done, or of taking steps automatically when no user intervention is needed.

One way to make GIS easier to use would be to raise the level of interaction to that of the conceptual data model--the continuous field, class of discrete objects, or linear network--rather than the level of implementation as is currently done. One uniform command language could be applied to all six representations of continuous surfaces, for example, hiding the details of the internal representation except when it is necessary that the user be aware of them. This would effectively reduce the number of data models currently implemented to three rather than at least ten, and would greatly reduce the complexity of current command languages and scripts.

For example, consider a simple request to a GIS to compute the mean rainfall for each U.S. state. Rainfall is conceived as a field, with a single value of the variable at each point in the U.S., and with the variable measured on a continuous (ratio) scale. Similarly "state" is conceived as a discrete or multinomial or nominal field, with a single value at each point in the U.S., measured on a scale which has only 50 values (51 if the District of Columbia is included) and in which value conveys no sense of order, ratio, or difference. At this point the query is fully defined, and there is no need for the user to supply any additional information for it to be resolved, assuming of course that the necessary data is available.

Current GIS vastly complicates this situation, by requiring its user to specify significantly more. We need to know whether the "state" coverage is raster or vector, and how the rainfall data is stored. Operations to convert raster to or from vector may have to be implemented, followed by the overlay of two layers, and a cross-tabulation. The associated script may stretch to ten or even 20 statements. But all of these operations can be anticipated, and none add to the specification of the original query. Parameters such as the cell size and numbers of rows and columns should be available in the associated metadata, though at this time there is very little implementation of metadata in current GIS. With a system such as this we would expect the results to include an estimate of the uncertainty associated with each value of rainfall, and again this should be computed without further intervention or specification by the user.

The research needed to allow this development to occur is largely complete, and experimental systems of this nature have already been built. Before they can be implemented, however, we will need to confront the kinds of issues raised earlier in connection with legacies--such systems will lack many of the familiar landmarks of GIS, including the function of "overlay", which is often regarded as the most important element of the GIS toolkit, but which is fully redundant under the scenario sketched here.

Expanding the domain

I argued earlier that data models ultimately constrain the degree to which software functionality can be integrated in any one package. Because of their historical roots, GIS data models emphasize the representation of maps and images, while word processing data models emphasize linear text, and spreadsheets focus on numeric tables. In this sense they have inherited many of the characteristics of paper maps. While it is easy to develop a full range of functions for this type of data, it is very difficult to extend a data model once it has been chosen as the basis of a package, and thus it has proven difficult for GIS to escape its heritage.

Specifically, paper maps are flat, requiring the use of complex map projections in order to represent the curved surface of the Earth. They are two-dimensional, and it has been difficult to adopt GIS technology to the representation of three-dimensional data, or objects embedded in three-dimensional space, except in cases where the third dimension can be regarded as a function of the other two (e.g., in representation of terrain). They are static, and it has been difficult to add the temporal dimension to GIS. Finally, they are of uniform scale, and it is not easy in GIS to link information across scales.

Early research focused on the representation of each of these extensions of the basic data model. Goodchild and Yang (1992), for example, have discussed structures for the sphere that have some of the characteristics of quadtrees; there is an extensive literature on true three-dimensional GIS (Turner 1992); Langran (1992) has discussed the implementation of time-dependence in GIS; and many solutions have been proposed for handling spatial hierarchies.

The next generation of GIS may take a different perspective on these issues, based on concepts of interoperability. Rather than a single, monolithic package, future GIS is likely to feature interoperable components. Thus three-dimensional capabilities may exist in one package, and may communicate with a traditional two-dimensional GIS by sending and receiving appropriately structured objects, using open object protocols, and through remote procedure calls. The Open GIS Consortium (http://www.ogis.org/overview.html) is developing the protocols to support this kind of interoperability.

In addition to these technological problems, many other issues combine to create impediments to any expansion of the domain of GIS. While it is easy to imagine a three-dimensional GIS, for example, we lack many of the essential ingredients to a three-dimensional GIS application. There is very little truly three-dimensional geographic data. We lack well-known, exemplary applications, except in limited areas. Finally, there are no familiar metaphors, such as the paper map, to guide our thinking about these extensions. In short, the paper map creates a powerful legacy.

Additional directions

The previous section identified four possible extensions to GIS data models, each of which has been the subject of recent research (see, for example, Molenaar and De Hoop 1994). But this analysis raises the question of whether other directions might also exist. The paper map has a uniform scale, and also a uniform quality of coverage, except where validity diagrams indicate variation in currency or other aspects of quality. Programs such as national topographic mapping have encouraged us to believe that uniformity of quality is a natural condition of geographic data. But in the future it seems unlikely that that assumption will remain true. The U.S. National Spatial Data Infrastructure (NSDI), for example, will be created through a series of partnerships, and its framework data will likely vary in scale, with large-scale data for urban areas and small-scale data for rural. The new satellite sources now coming on line acquire data on command rather than in fixed swaths, so their output is likely to be distributed in the form of mosaics rather than as fixed scenes. Global databases that have been compiled from national sources are also likely to show significant variation in quality.

At this time we have very little in the way of data models for data of varying uncertainty. In such circumstances metadata is likely to take the form of a map, rather than a set of records, more akin to the validity diagram of a topographic sheet than the card catalog. We lack analytic methods for data of variable quality, and have a poor level of understanding of the implications of varying quality for the results of analysis.

Platform interoperability

Finally, I suggest interoperability between platforms as a fourth area for GIS development. Earlier I discussed the potential for interoperability between the six distinct field data models, since each implements the same basic conceptual model of a continuous field. In principle, we ought to be able to extend this argument to the case of data models operating in different computing environments. For example, it ought to be possible to resolve the query "determine the average rainfall by state in the U.S." even though rainfall is stored in an ARC/INFO coverage on a Unix platform, and the state coverage is in IDRISI on a PC running under Windows 95. The same argument applies--the extensive script that would need to be written in today's GIS environment adds nothing to the specification of the original query that could not be resolved automatically by the computing environment without user intervention.

This example serves to emphasize once again the essential role played by metadata in achieving interoperability within the data life cycle. Metadata allows data to describe itself to other systems, in addition to allowing us to find data, share it, and assess its fitness for use. It adds value to data by making it possible for others to use it, thus ensuring better return on our original investment.

Concluding comments

As I noted at the outset, this view of GIS research directions is strictly my own, and I have undoubtedly omitted mention of many interesting research areas that are equally important to the future of the field. The pages of the journals, notably the *International Journal of Geographical Information Systems*, and the proceedings of recent conferences give a much more complete picture.

In this paper I have tried to stress several general and interlocking themes. One is the influence of new thinking, and its relationship to the development of technology--the developments in digital technology that have occurred over the past few years have stimulated an enormous amount of new thinking on the future, and the likely form of the advanced technological society that is now emerging. Another is the transition of GIS from a technology of analysis to one that underlies and facilitates the entire data life cycle, including data collection and library search. A third is the need for GIS to become easier to use, and easier to integrate with other technologies.

Acknowledgment

The National Center for Geographic Information and Analysis and the Alexandria Digital Library project are supported by the National Science Foundation.

References

Cowen, D.J. (1988) GIS versus CAD versus DBMS: what are the differences? *Photogrammetric Engineering and Remote Sensing* 54: 1551-1554.

Egenhofer, M.J., and J.R. Herring, editors (1995) *Advances in Spatial Databases: Fourth International Symposium, SSD '95, Portland, Maine, August 6-9, 1995*. Lecture Notes in Computer Science 951. Berlin: Springer-Verlag.

Frank, A.U., and W. Kuhn, editors (1995) *Spatial Information Theory: A Theoretical Basis for GIS: International Conference, COSIT '95, Semmering, Austria, September 21-23, 1995*. Lecture Notes in Computer Science 988. Berlin: Springer.

Goodchild, M.F. (1987) A spatial analytical perspective on GIS. *International Journal of Geographical Information Systems* 1(4): 327-334.

Goodchild, M.F. (1992) Geographical data modeling. *Computers and Geosciences* 18(4): 401-408.

Goodchild, M.F., and S.R. Yang (1992) A hierarchical spatial data structure for global geographic information systems. *CVGIP-Graphical Models and Image Processing* 54(1): 31-44.

Langran, G. (1992) *Time in Geographic Information Systems*. London: Taylor and Francis.

Maguire, D.J. (1991) An overview and definition of GIS. In D.J. Maguire, M.F. Goodchild, and D.W. Rhind, editors, *Geographical Information Systems: Principles and Applications*, Vol 1, pp. 9-20. London: Longman Scientific and Technical.

Maguire, D.J., and J. Dangermond (1991) The functionality of GIS. In D.J. Maguire, M.F. Goodchild, and D.W. Rhind, editors, *Geographical Information Systems: Principles and Applications*, Vol 1, pp. 319-335. London: Longman Scientific and Technical.

Molenaar, M., and S. De Hoop, editors (1994) *Advanced Geographic Data Modelling: Spatial Data Modelling and Query Languages for 2D and 3D Applications*. Publications on Geodesy, New Series, No. 40. Delft: Netherlands Geodetic Commission.

Novak, K., and J.D. Bossler (1995) Development and application of the highway mapping system of Ohio State University. *Photogrammetric Record* 15(85): 123-134.

Tomlinson, R.F. (1967) *An Introduction to the Geographic Information System of the Canada*

Land Inventory. Ottawa: Department of Forestry and Rural Development.

Turner, A.K. (1992) *Three-Dimensional Modeling with Geoscientific Information Systems*. Dordrecht: Kluwer.

Waugh, T.C., and R.G. Healey, editors (1994) *Advances in GIS Research: Proceedings of the Sixth International Symposium on Spatial Data Handling*. London: Taylor and Francis.

Michael F. Goodchild
Director, National Center for Geographic Information and Analysis
Professor, Department of Geography
University of California
Santa Barbara, CA 93106-4060, USA
+1 805 893 8049
FAX +1 805 893 7095
good@ncgia.ucsb.edu

Jochen H. Albrecht

Universal GIS Operations for Environmental Modeling

In the first part of this paper, the author presents a set of 20 universal analytical GIS operations, i.e. operations that are independent of data structures, yet cover the full range of analytical capabilities of current vendor-based GIS. This set represents a user's task-oriented view of GIS functionality rather than that of a technically oriented developer. Their function is readily apprehended by any spatially aware individual and does not require any knowledge about abstract spatial concepts. These operations constitute the foundation of a shell that has been developed at the Institute for Spatial Analysis and Planning in Areas of Intensive Agriculture (ISPA) and goes by the name Virtual GIS (VGIS). VGIS provides an ideal tool box for the environmental modeler, who wants to concentrate on modeling issues rather than the intricacies of GIS.

In the second part, the flow chart-based user interface of VGIS is introduced as a visual programming and prototyping tool similar to STELLA® (HPS, 1994), but working with real GIS data and offering the full functionality of GIS. The conceptual modeling capabilities are exemplified by applications from environmental modeling.

---

1. INTRODUCTION

The domain of universal GIS methods is an area of Geographic Information Science (Goodchild, 1992a) that so far has been neglected by research and for which therefore no body of theory exists. Driven by the heterogeneous market forces every vendor and a multitude of academic developers produced myriads of commands to perform GIS operations. The few existing taxonomies of GIS operations (Aronoff, 1991; Burrough, 1989 and 1992; Goodchild, 1992b; de Man, 1988; Rhind and Green, 1988; Unwin, 1990; Schenkelaars, 1994) are limited either by the data structure that they are based on or by the scope of applications for which they had been developed. They all lack formalization and do not attempt to be truly universal. Section 2 introduces the author's work on a task-oriented systematization of data structure-independent GIS functionality.

On the ecological modeling side, Constanza and Sklar (1985) state that "most ecological modeling work to date has focused on temporal changes. They tend to simulate a point in space and extrapolate the findings for an entire landscape by assuming that the landscape is *homogeneous*. In other words, most models in ecology have little, if any, spatial articulation. It is clear, however, that space needs to be more explicitly included if ecological models are to be truly useful tools for understanding and predicting the behavior of real ecosystems (Risser, *et al.*, 1984)." Although this quote is 10 years old now, it has not lost any of its truth. Section 3 reviews the current state in GIS-based ecological modeling.

Section 4 tries then to apply the universal GIS operations described in section 2 in an environmental modeling context. Special emphasis is given to the phase of conceptual modeling as well as to an object-oriented application of individual modeling units. The conclusion (section 5) describes current limitations to the proto-type and discusses some rather fundamental ideas about the synthesis of GIS and environmental modeling.

2. UNIVERSAL GIS OPERATIONS

One motivation for research on universal GIS operations simplifying that simplify GIS use was the observation, that in spite of its name, current GIS have little to offer to the scientist, who is interested in modeling spatial phenomena. Tomlin's (1990) cartographic modeling language (also known as 'Map Algebra') is the most sophisticated GIS modeling environment so far. This lack has been articulated and mourned by many in the modeling community and resulted in a conference series, devoted to overcome this discrepancy between the GIS and the modeling community (Goodchild *et al*., 1993; Goodchild *et al*., 1995).

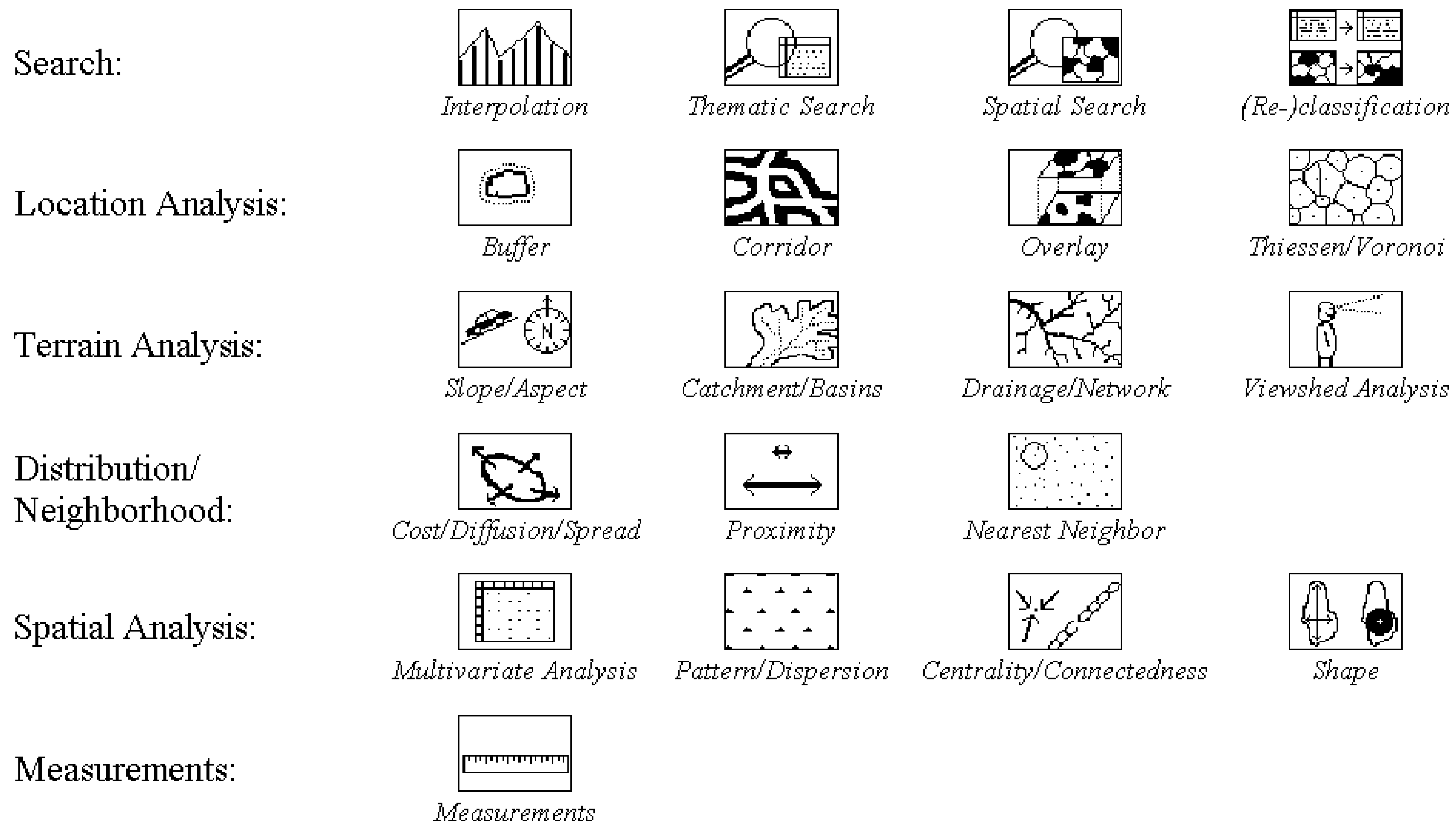
Current Geographic Information Systems (GIS) are so difficult to use that it takes some expertise to handle them, and it is not unusual to assess a whole year until an operator masters a GIS. This is especially cumbersome for cursory users (such as environmental modelers) that employ GIS as one tool among many others. Coulsen, *et al.*, (1991) expressed a similar argument when they wrote "GISs are complex computer programs. Proficient application in natural resource management and landscape ecology involves a commitment to training and practice by the user. None of the GISs would be considered *user friendly* by a human factors engineer." Similar comments may be found throughout (Medyckyj-Scott and Hearnshaw, 1993) and (Turk, 1992).

Although a number of GIS claim to be data structure-independent none of them really is; they all show their origin as so-called either raster (grid cell-based) or vector systems. This data structure distinction has dictated differences in analysis functionality. Even Goodchild (1991a, p. 45) in his often cited classification of spatial data analysis techniques groups them depending on the underlying data model. And while there are numerous efforts to standardize data models (SAIF, SDTS, DIGEST, GDF), so far none of these attempt to standardize the operations as well. The advent of the Open Geodata Interoperability Specification (OGIS) (Buehler, 1994; Buehler, 1995) opens for the first time a real opportunity to develop data structure-independent GIS applications (see also the contribution by Kenn Gardels in this volume). As far as is known so far, however, the Open GIS Consortium (OGC) does not attempt to define high level operations but restricts its specification to low level database (SQL-like) and topological operations based on the work by Egenhofer (1991, 1993).

The author tackled these deficiencies with a user survey to determine *user* expectations of a GIS' functionality (Albrecht, 1995b). The responses reveal a vast array of complexity ranging from elementary operations to compound tasks. A dissection of the latter into fundamental primitives leads towards a normalization of GIS operations. Since the analytic capabilities of GIS are the only ones that distinguishes them from other visualization software or database management systems (Burrough, 1986; Goodchild, 1987), further consideration of GIS functionality within this paper will concentrate on this ability.

By analyzing current GIS user interfaces and omitting all those operations that are due to either the historic development of the particular software package or are a result of the data model employed, the author derived a list of only 20 universal GIS operations that allow to build all but the most exotic GIS applications (see also Table 1). This small set of spatial analytical tools provides the means to perform environmental spatial modeling without having to learn about the intricacies of current GIS. A detailed description of how this particular set of GIS operations was derived is given in (Albrecht, 1996). There, the reader will also find a formalization of these operations based upon a simplified version of the OGIS data model. An implementation of these fundamental GIS operations within an interactive flow-charting environment (see section 4) reveals the window of opportunities opened by this approach.

Table 1. The 20 universal GIS operations



Search:	*Interpolation*	*Thematic Search*	*Spatial Search*	*(Re-)classifi cation*
Location Analysis:	*Buffer*	*Corridor*	*Overlay*	*Thiessen/Voro*

				noi
Terrain Analysis:	*Slope/Aspect*	*Catchment/Basins*	*Drainage/Network*	*Viewshed Analysis*
Distribution/ Neighborhood:	*Cost/Diffusion/ Spread*	*Proximity*	*Nearest Neighbor*	
Spatial Analysis:	*Multivariate Analysis*	*Pattern/Dispersion*	*Centrality/Connectedness*	*Shape*
Measurements:	*Measurements*			

3. THE STATE OF ART IN THE APPLICATION GIS FOR ENVIRONMENTAL MODELING

> *"GIS can be used to seduce the user into an unrealistic sense of model accuracy", using "a few poor, anemic point measurements".*
> (Grayson, *et al.*, 1993, p. 91)

One motivation for the search for the GIS usage simplifying universal GIS operations was the observation, that in spite of its name, current GIS have little to offer to the scientist, who is interested in modeling spatial phenomena. Tomlin's (1990) cartographic modeling language (also known as 'Map Algebra') is the most sophisticated GIS modeling environment so far. This lack has been articulated and mourned by many in the modeling community and resulted in a conference series, devoted to overcome this discrepancy between the GIS and the modeling community (Goodchild, *et al.*, 1993 and Goodchild, *et al.*, 1995).

The two probably most exhaustive overviews (Sklar and Constanza, 1991; Hunsaker, *et al.*, 1993) describe numerous environmental tasks such as inventory, assessment, management, and predicting the fate of environmental resources supporting applications in atmospheric modeling, hydrological modeling, land surface-subsurface modeling, ecological systems modeling, plus integrated environmental models as well as policy considerations for risk/hazard assessment involving these models (see Table 2). All of these, however, use a GIS as an inventory for spatially referenced data and for presentation (map production) only.

Nyerges (1991) identified three primary modes of GIS use, namely *map mode* providing referential and browse information, *query mode* to address specific requests for information based either on field or object views, and *model invocation* which is the only mode that makes use of the analytical capabilities of a GIS. On a rather abstract level Nyerges describes the development of a typical modeling process, however, he fails to note that up to now there exists no GIS that actually supports such a procedure.

Two research projects focus on the problem of facilitating a computational modeling environment employing rather different approaches. The first one to be described here is the computational modeling system (CMS) developed at the Department of Computer Science at the University of California (Smith, *et al.*, 1995). Their computational modeling language (CML) is supposed to support cooperative (geographic) modeling activities at all stages of the modeling process, i.e. data extraction, construction and evaluation of conceptual models, model refinement, and communication of the results (Alonso and Abbadi, 1994). The CML is based upon the concept of so-called representational (or -) structures and their transformations. These -structures contain specifications how to represent the same information using a different data model, so the user does not need to explicitly know about the data model that the source data is based on. Each -structures then also contains the operations that can be applied on it, e.g. a digital elevation model (DEM) may contain the transformations '*union*' to combine several DEMs, '*compute-slope*', or '*max-height*'. Creating such schemas nevertheless requires to learn a new programming language which might be worth the effort for a some model builders, but renders it unlikely that CMS will become a wide-spread tool.

The other research project aimed at facilitating modeling procedures within a GIS is the Virtual GIS (VGIS) project described in Albrecht (1995a). VGIS is a shell that employs the universal GIS operations described in section 2 using a flow charting environment (see Figure 1). Flow charts are a the standard process-oriented tool in visual programming (Chang, 1990; Glinert, 1988; Monmonnier, 1989). As in the case of CMS, the system has yet to prove its usability in real world (meaning non-academic) applications. Proposed (and already granted, see section 6) extensions of this work promise a wider applicability, as one of the interpreters that is intended to be developed for VGIS will be an interface to the OGIS data model.

Table 2. Spatial Models in Different Domains (after Sklar and Constanza, 1991)

Geography

geometric (all descriptive; without feedback loops)

von Thünen (space around a point)

Horton's Law

Christaller (hierarchy of hexagons)

demographic

expansive/relocative

Hägerstrand's innovation diffusion (logistic curve)

Diffusion modeled as a Markovian process using gravity models $T=kW/d^a$ with k and a being empirical measures or spatial stochastic parameters;

i.e. autocorrelation

network (linear programming; assumes linear relationships; no feedbacks)

applications: traveling salesman, commodity flow, spatial allocation, optimization

Hydrology (physical continuity equations)

finite element (hydrodynamic)

link-node design similar to network models in geography

dynamics are a function of the network design and the simultaneous solution of the continuity and momentum equations of every node

no spatial processes in non-link areas, i.e. adjacent to pipelines

solutions to the latter problem:

link size as function of the area surrounding adjacent nodes

parallel pipes

areas adjacent to pipes become reservoirs

finite difference (general circulation)

General Atmospheric Circulation Models (GACM)

3-dimensional mesh; grid along latitude/longitude, 10 or 100 layers

Biology

growth (landscape is an independent exogenic variable)

competitive interaction, i.e. Lotka-Volterra

patch dynamics

interference and exploitation

niche space

population

very similar to geographic demography with more emphasis on density-dependence
in plant ecology: correlation between abundance and spatial heterogeneity;

they emphasize community-level rather than population-level interactions

point-averaged ecosystem

STELLA®-like; condense complex ecosystems into a small number of differential equations

Ecosystem Research (Landscape)

stochastic

fit the definition of spatial modeling (Griffith and Mac Kinnan, 1981)

map the flow of energy, matter and information, designate source, sink and
receptor areas, predict succession in 2D and 3D space, address questions of scale

incorporation of feedback loops, neighborhood influences and spatial im-/exports

results in spatially more articulate models

process-based

analog to the Global Atmospheric Circulation Models above;

grid-type, 2D or 3D flow, material transport as a function of mass balance

only process-based landscape models have the advantage of being spatially explicit, realistic and dynamic

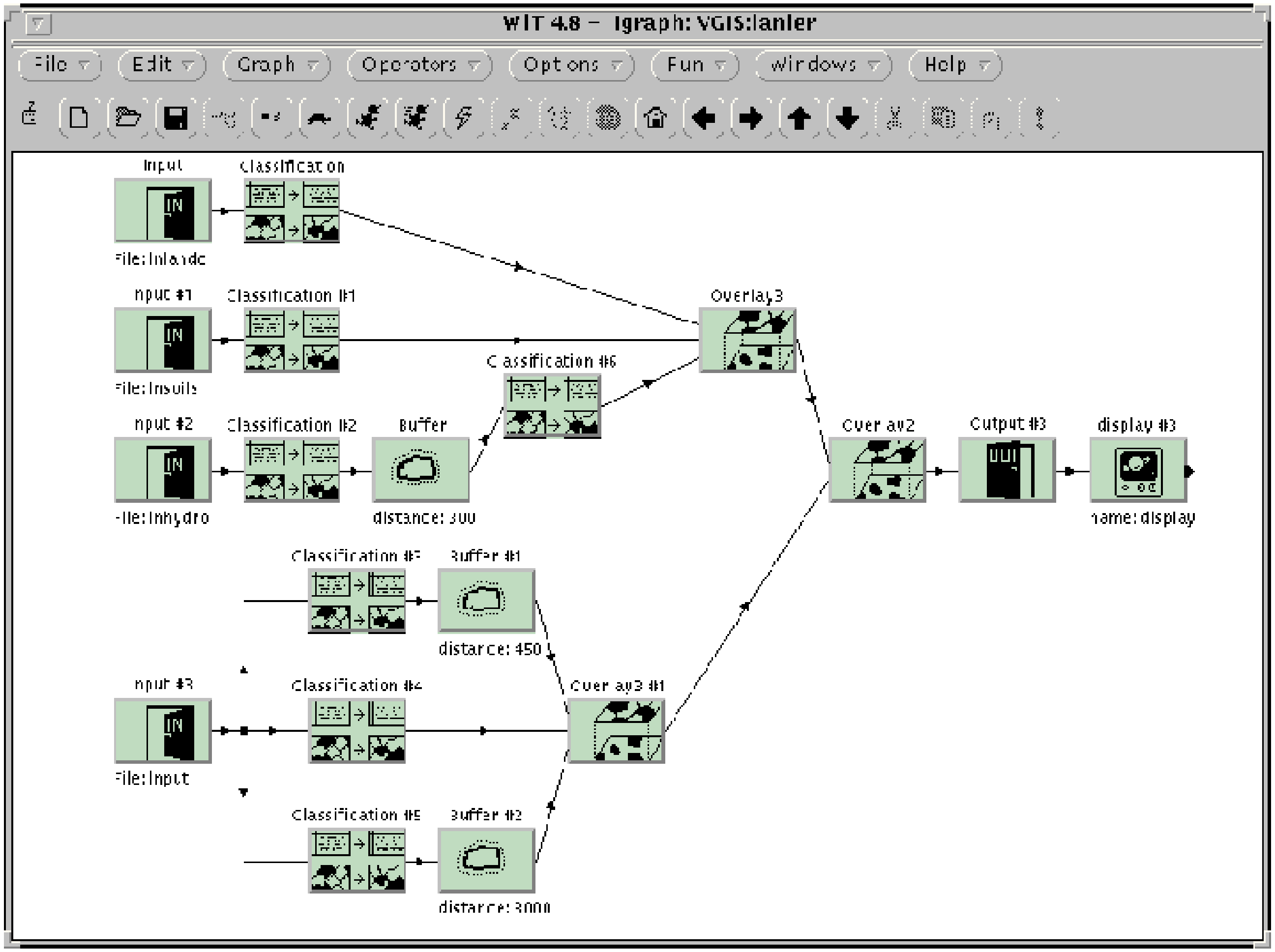


Figure 1. VGIS modeling example (screen dump)

4. ENVIRONMENTAL MODELING WITH VGIS

So far, integrated application of GIS and environmental modeling have been specialized modeling languages, such as PCRaster and DYNAMO in the hydrologic domain (van Deursen and Kwadijk, 1993; Wesseling and v. Deursen, 1995). These languages, however, do not really support the creative process of model building. Rather, they require an intricate knowledge of the model and the language, and are harnessed to fine-tune a fixed model run. The Virtual GIS (VGIS) on the other hand, attempts to be a prototyping tool and development platform similar to STELLA® (HPS, 1994), but working with real GIS data and thereby graphically extending 'Map Algebra' according to the concepts presented in (Kirby and Pazner, 1990). More realistic (real world) data with a locational character have a significant impact on the model results. In addition, geographical displays interactively depicting the nature of the sensitivity of certain parameters can be useful in support of model parameterization. examining scale effects can be accomplished by changing interactively the nature of the data aggregation. The model brings together the locational, temporal, and thematic aspects of phenomena in a geographic process characterization.

Such a visual programming example is depicted in Figure 2, where the modeling flow chart allows the user to "play" with the data flow. Figure 2 represents an intermediate step in the conceptual modeling of erosion. The four input files (rounded boxes) are geology, landcover, precipitation, and elevation. It is possible to model this complex system with only five of the universal GIS operations. Within VGIS, it is easy to test the result of new routing paths within the flow chart. The hypothesis that a certain region buffered around drainage channels has a different water retention capacity can easily be tested by adding one connection to the flow chart. A similar reconfiguration of a conceptual model would require substantial GIS expertise if it were attempted in a vendor GIS. The similarity with the previously mentioned prototyping tool extends to another feature as well: each data object might explode into another model that can either be treated as a black box or specified in a similar manner as its parent level.

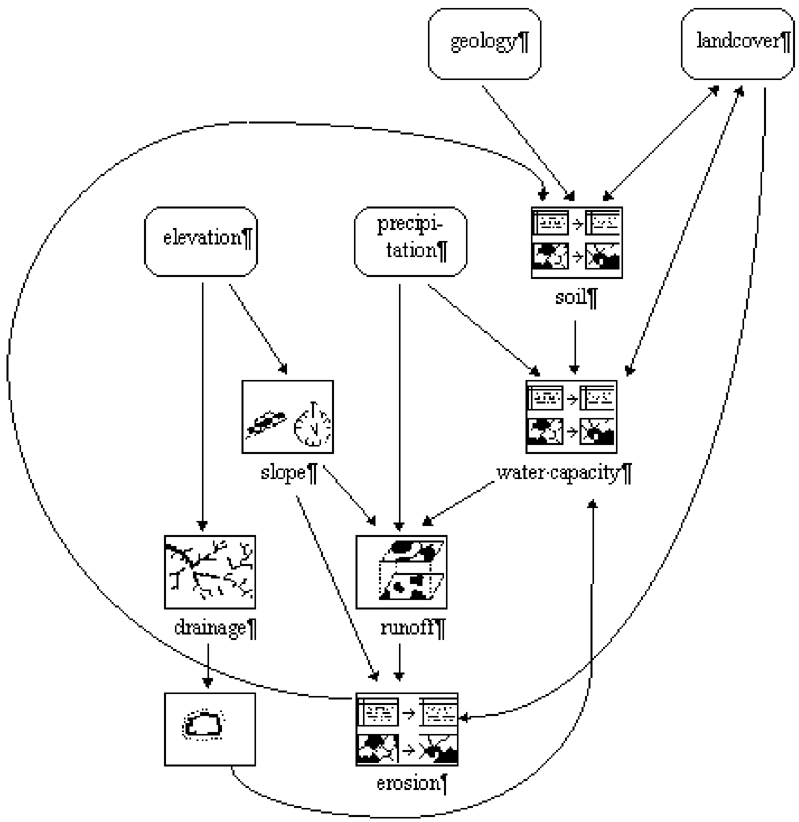


Figure 2. Sample modeling flow chart

Probably the main advantage of the VGIS environment lies in the first time-ever possibility to easily include feedback loops within a GIS. "Landscapes are never static; their elements are in constant temporal and spatial flux" (Merriam, *et al.*, 1991). Sklar and Constanza (1991) therefore consider the incorporation of space as well as time as the most prominent issue in ecosystem research. This needs to be done at all levels of resolution that are meaningful to the myriad ecosystem management problems we now face. It is this explicitly spatial aspect is what motivates landscape ecology.

Another major advantage of the VGIS environment is the inclusion of spatial statistics as GIS functionality. "To consider something a system, it is necessary to describe its boundary and its interaction with the environment" (Frank, *et al.*, 1994). Therefore, one of the first tasks within a modeling environment should be to delineate the spatial boundaries. This can be readily accomplished by calling the '*Pattern/Dispersion*' analysis operation. Although this is one of the core functions of a GIS, the author has yet to come across a reference for an environmental modeling application where a GIS is actually used for this purpose.

5. CONCLUSION

VGIS is implemented as an interpreter to the public domain GIS GRASS. Current work focuses on the development of another interpreter for Arc/Info®. The only handicap for a free distribution of the prototype is the utilization of the flow charting tool WiT®. A detailed technical description is given in (Brösamle, *et al.*, 1996). The VGIS environment does not yet include conditional and iteration operators as they are used in formal programming languages. Therefore, in its current state, the VGIS can not be called a geographic modeling language yet.

One of the main impediments to the application of VGIS in real world applications has been the difficulty to environmental modeling tasks that require the full scope of analytical GIS functionality. Most examples would do well with a minimal set of 'overlay', 'slope calculation' and 'diffusion/spread'. The latter two are often implemented in cell-based modeling systems as well. This experience comes very much as a surprise to a geographer who started out with conviction that space is such an important feature in any environmental model that the full power of GIS will be required to satisfy even minimal scientific deeds. Now the far more modest author recognizes that there is still a long way to go, before we will understand the rules governing landscape ecology before we can implement a modeling system on a truly large scale.

6. ACKNOWLEDGEMENTS

This research is supported by the German Science Foundation which provides the funding for the Virtual GIS (VGIS) project under grant number IIC5-Eh 85/3-2.

7. REFERENCES

Albrecht, J., 1995a. Virtual Geographic Information System (VGIS). In Nunamake, J. and R. Sprague (Eds): *Proceedings 28th Hawaii International Conference on System Sciences*, GIS Minitrack, Volume IV, pp. 141-150. IEEE Computer Society Press, Los Alamitos, CA.

Albrecht, J., 1995b. Semantic Net of Universal Elementary GIS Functions. *Proceedings ACSM/ASPRS Annual Convention and Exposition Technical Papers*, Vol. 4 (Auto-Carto 12), pp. 235-244. Bethesda, MD.

Albrecht, J., 1996. *Universal GIS Operations*. Dissertation, published electronically and accessible as http://www.ncgia.ucsb.edu/~jochen/diss/dissabst.html.

Alonso, G. and A. Abbadi, 1994. Cooperative Modeling in Applied Geographic Research. *International Journal of Intelligent and Cooperative Information Systems*, 3(1):83-102.

Aronoff, S., 1991. *Geographic Information Systems: A Management Perspective*. WDL Publications, Ottawa, 2°, 294 pp.

Brösamle, H., Albrecht, J. and M. Ehlers, 1996. GIS Functionality and Interface Design: The Virtual GIS (VGIS) Example. Invited paper, ISPRS Commission II Working Group II/2, Workshop on New Developments in Geographic Information Systems, Milan, Italy.

Buehler, K., 1994. OGIS Project Document 94-019. *OGIS Geodata Model Overview*. The Open GIS Foundation, Cambridge, MA.

Buehler, K., 1995. Open Geodata Interoperability Specification (OGIS). Presentation given by Kurt Buehler at the OGIS workshop in Zürich, Switzerland, 30 November 1995.

Burrough, P., 1986. *Principles of Geographic Information Systems for Land Resources Assessment.* Monograph on Soil and Resources Survey, No. 12. Clarendon Press, Oxford. 194 pp.

Burrough, P., 1992. Development of Intelligent Geographical Information Systems. *Internatinoal Journal of Geographic Information Systems*, 6(1):1-11.

Chang, S., 1990. *Principles of Visual Programming Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Constanza, Sklar, 1985. Articulation, Accuracy, and Effectiveness of Mathematical Models: a review of freshwater wetlands applications. *Ecological Modeling*, 27(1):45-68.

Coulsen, R., Lovelady, C., Flamm, R., Spradling, S. and M. Saunders, 1991. Intelligent Geographic Information Systems for Natural Resource Management. In Turner, M. and R. Gardner (Eds.): Quantitative Methods in Landscape Ecology, pp. 153-172, New York, Springer-Verlag.

de Man, E., 1988. Establishing a geographic Information System in Relation to its Use. *International Journal of Geographic Information Systems*, 2(3):257.

Egenhofer, M. and R. Franzosa, 1991. Point-Set Topological Spatial Relations. *International Journal of Geographical Information* Systems 5(2):161-174.

Egenhofer, M., Sharma, J. and D. Mark, 1993. A Critical Comparison of the 4-Intersection and the 9-Intersection Models for Spatial Relations: Formal Analysis. Proceedings, *Eleventh International Symposium on Computer-Assisted Cartography* (Auto Carto 11), pp. 1-11.

Frank, A., Egenhofer, M. and D. Hudson, 1994. The Design of Spatial Information Systems. Part 1: Formal Systems. File on ftp://grouse.spatial.maine.edu/pub/SurveyEng/sve451/451.ps.

Glinert, E. (Ed.), 1988. Visual Programming Environments: applications and issues. IEEE Computer Society Press, Los Alamitos.

Goodchild, M., 1991. Spatial Analysis with GIS: problems and prospects. *Proceedings GIS/LIS'91*, pp. 41-48. ASPRS/ACSM, Falls Church, VA.

Goodchild, M., 1992a. Geographical Information Science. *International Journal of Geographic Information Systems*, 6(1):31-46.

Goodchild, M., 1992b. *Spatial Analysis Using GIS*: a seminar workbook, pp. 40-48, 2°, National Center for Geographic Information and Analysis, University of California, Santa Barbara.

Goodchild, M., Parks, B. and L. Steyaert (Eds.), 1993. Environmental Modeling with GIS. Oxford University Press, New York/Oxford.

Goodchild, M., Steyaert, L., Parks, L., Crane, M., Johnston, C., Maidment, D. and S. Glendenning, 1995. GIS and Environmental Modeling: Progress and Research Issues. GIS World, Inc., Ft. Collins, CO.

Grayson, R., Blöschl, G., Barling, R. and I. Moore, 1993. Process, Scale and Constraints to Hydrological Modelling in GIS. In Kovar, K. and H. Nachtnebel (Eds.) *Application of Geographic InformationSystems in Hydrology and Water Resources Management*. Proceedings HydroGIS 93. Vol. 211, pp. 83-92, IAHS Press, Wallingford, England.

High Performance Systems (HPS), 1994. STELLA II: an introduction to systems thinking. Hanover, NH.

Hunsaker, C., Nisbet, R., Lam, D., Browder, J., Baker, W., Turner, M. and D. Botkin, 1993. *Spatial Models of Ecological Systems and Processes: The Role of GIS*. In Goodchild *et al.*, (1993), pp. 248-264.

Kirby, K. and M. Pazner, 1990. Graphic Map Algebra. In Brassel/Kishimoto (Eds.): *Proceedings of the 4th International Symposium on Spatial Data Handling*, Vol. 1, pp. 413-422.

Medyckyj-Scott, D. and H. Hearnshaw, 1993. *Human Factors in Geographic Information Systems*. Belhaven Press, London, 266 pp.

Merriam, G, Henein, K. and K Stuart-Smith, 1991. Landscape Dynamics Models. In Turner, M. and R. Gardner (Eds.): *Quantitative Methods in Landscape Ecology*, pp. 399-416, Springer, New York, NY.

Monmonnier, M., 1989. Graphic Scripts for the Sequenced Visualization of Geographic Data. *Proceedings GIS/LIS'89*, pp. 381-389. ASPRS/ACSM, Falls Church, VA.

Nyerges, T., 1991. Analytical Map Use. *Cartography and Geographic Information Systems*, 18(1):11-22.

Rhind, D. and N. Green, 1988. Design of a Geographical Information System for a Heterogeneous Scientific Community. *International Journal of Geographic Information Systems*, 2(2):175.

Risser, P., Karr, J. and R. Forman, 1984. *Landscape Ecology: Directions and Approaches*. Special Publication No. 2. Illinois Natural History Survey, Champaign, IL (qf. Sklar and Constanza, 1991).

Schenkelaars, V., 1994. Query Classification, a First Step Towards a Graphical Interaction Language. In Molenaar, M. and S. de Hoop, (Eds.): *Advanced Geographic Data Modeling*. Netherlands Geodetic Commission, Vol. 40, pp. 53-65.

Sklar, F. and R. Constanza, 1991. The Development of Dynamic Spatial Models for Landscape Ecology: a review and prognosis. In Turner, M. and R. Gardner (Eds.): *Quantitative Methods in Landscape Ecology*, pp. 239-288, Springer, New York, NY.

Smith, T., Su, J., Abbadi, A., Agrawal, D., Alonso, G. and A. Saran, 1995. Computational Modeling Systems. *Information Systems*, 20(2):127-153.

Tomlin, D., 1990. *Geographic Information Systems and Cartographic Modeling*. Prentice-Hall, Englewood Cliffs, NJ.

Turk, A., 1992. *GIS Cogency: cognitive ergonomics in geographic information systems*. Unpublished dissertation, Department of Surveying and Land Information, University of Melbourne.

Unwin, D., 1990. A Syllabus for Teaching Geographical Information Systems. *International Journal of Geographic Information Systems*, 4(4):461-462.

van Deursen, W./Kwadijk, J., 1993. RHINEFLOW: an Integrated GIS Water Balance Model for the River Rhine. In *Applications for Geographic Information Systems in Hydrology and Water resources*. IAHS Publications No. 211, pp. 507-518.

Wesseling C., and W. van Deursen, 1995. A Spatial Modelling Language for Integrating Dynamic Environmental Simulations in GIS. *Proceedings Joint European Conference and Exhibition on Geographical Information*, Vol. 1, pp. 368-373.

Jochen H. Albrecht
Institute for Structural Analysis and Planning in Areas of Intensive Agriculture (ISPA)
University of Vechta
Postfach 1553
49364 Vechta, Germany
Telephone: +49 (444) 11 53 37
FAX: +49 (444) 11 54 45
http://www.geog.ucsb.edu/~jochen/

Virtual Data Sets - Smart Data for Environmental Applications

Andrej Vckovski and Felix Bucher

Spatial Data Handling
Department of Geography
University of Zurich
Winterthurerstr. 190
CH-8057 Zürich
Switzerland
vckovski@geo.unizh.ch / bucher@geo.unizh.ch

Abstract

Continuous fields on a spatio-temporal support are one of the basic types of data used in environmental modeling, e.g., air temperature and pressure, wind-fields, precipitation, soil types etc. The sampling of a continuous field necessarily involves discretization both in spatial and temporal domain. Data describing (random) continuous fields therefore consist of a series of samples at fixed spatial and temporal locations, e.g. as point values or as aggregations over certain areas and time intervals. For many applications the representation available with such data sets does not meet the requirements of the application. It is therefore often necessary to model the field under consideration based on the available data values and then to use the model to predict values at unsampled locations, i.e., to generate a new representation of the field.

The process of creating new representations, e.g., resampling or interpolation, is often time consuming and has strong impacts on the quality of the generated representation. Virtual Data Sets (VDS) are an approach to address these problems and improve reliability and re-usability of field representations. The concept is based on an extension of a data set with methods that implement a model of the field under consideration. That is, a VDS contains itself methods to generate new representations or to present itself as a new representation, respectively.

The structure of VDS leads to an implementation design based on a distributed computing environment and object-oriented technology. Therefore, VDS are implemented as distributed objects or services which can be queried by applications needing data. This design is discussed in the context of the sapproach taken by the Open Geodata Interoperability Specification (OGIS). The last part of this contribution discusses a possible implementation strategy based on Java. It is shown that Java can be used to realize a distributed geoprocessing environment and Virtual Data Sets.

Introduction

Environmental information is increasingly required for policy making in diverse fields such as natural risk assessment, crop yield estimation and ressource management. Such decision support information is provided by reliably analyzed environmental data, i.e., *value added* information. Over the past few years, Geographical Information Systems (GIS) have been evaluated as (and extended towards) an appropriate technology for environmental data management and analysis. This is mainly due to the fact that GIS has the ability to *integrate* diverse information from various sources which is a major requirement in environmental data analysis. Environmental data analysts, however, are still facing a whole bunch of problems when using GIS:

- Data acquisition is often the most expensive component of GIS-based environmental analysis. As a consequence, the re-use of available data should be maximised. Data exchange needs to be promoted. Tools for finding and acquiring suitable data sets are needed.
- Data integration needs not only standardised data exchange facilities but also techniques for overcoming higher level incompatibilities such as semantic integration (Bucher *et. al.*, 1994).
- Environmental data management and analysis is a demanding application for GIS using a wide range of analysis methods and data management techniques (Albrecht, 1995). These requirements resulted in the large and monolithic commercial GIS currently available. A more flexible and more powerful approach of GIS technology is based on modular and component-based systems such as proposed by OGIS (Buehler, 1994).

As a consequence and in order to overcome these problems, the GIS industry is now adopting new trends and technologies from fields such as computer sciences and data engineering. More specifically, GIS are moving towards true distributed, component-based geoproccessing across large networks as well as geodata interoperability. Standardization efforts based on object-oriented geodata models find increasingly acceptance in the design of modern GIS.

This paper presents the use of Virtual Data Sets (VDS) for environmental data management and analysis. More specifically, VDS provides a promising means for the storage and usage of data describing *continuous fields* as is often required in environmental applications. The VDS approach is based on "smart" data sets and views data sets as active objects in a distributed software system.

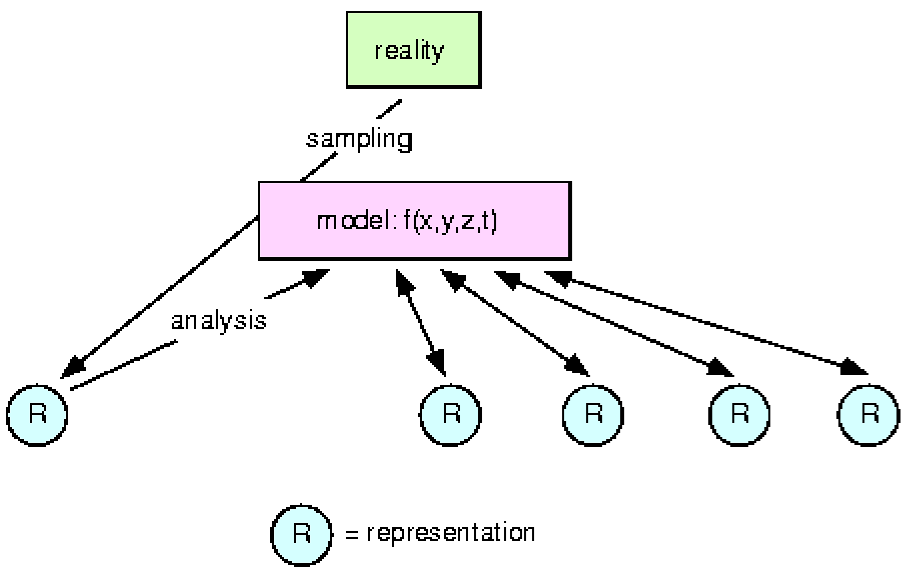
Interoperability Issues in Environmental Applications

The following discussion will be focused on *(continuous) fields*. Fields are the major information type used in most environmental applications. A field is basically a function relating the elements of a definition domain (support) to a value domain. The support is in most cases a subset of (physical) space and time, e.g., a time interval and/or a region on the earth's surface. The values of a field are either discrete or continuous, scalar- or vector-entities (Vckovski, 1995, Bucher & Vckovski, 1995). If the support is dense and compact then a field is called *continuous*.

The sampling of a field necessarily implies a discretization of the support, i.e., the field values are measured at a *specific set of sites* (sample sites). These sites are in most cases points (e.g., meteorological measurement devices) or simple polygons (e.g., squares on a satellite image). A *representation* of a field therefore consists of a finite set of sites and corresponding field values. Data sets typically are collected for a specific application in mind, with a corresponding sampling strategy. Due to the high costs of sampling, the acquired data needs to be re-used as much as possible. The *original* field representation, however, often does not meet the data requirements of future applications, i.e., the representations are semantically incompatible. Therefore, new representations of the original field representation have to be derived. Consider a digital elevation model (DEM) that was originally sampled at irregularly distributed points. In order to meet the requirements of some applications, the DEM - describing the terrain height field - might be transformed into other representations, such as:

- Height values at *other locations* (e.g., a regular grid).
- Slope and aspect of the terrain height by calculating the *gradient* field.

These transformations necessarily need a *model* of the field based on the sampled values. The model is a simplified view of the real-world behaviour of the underlying phenomenon. The transformation generates a *new* representation *through* the model. This is shown in the following figure:



The sampling (A) generates the original field representation. This representation then is *analyzed* to form a model which serves for the generation of another representation.

The process of a transformation is a critical task and bears many difficulties in it:

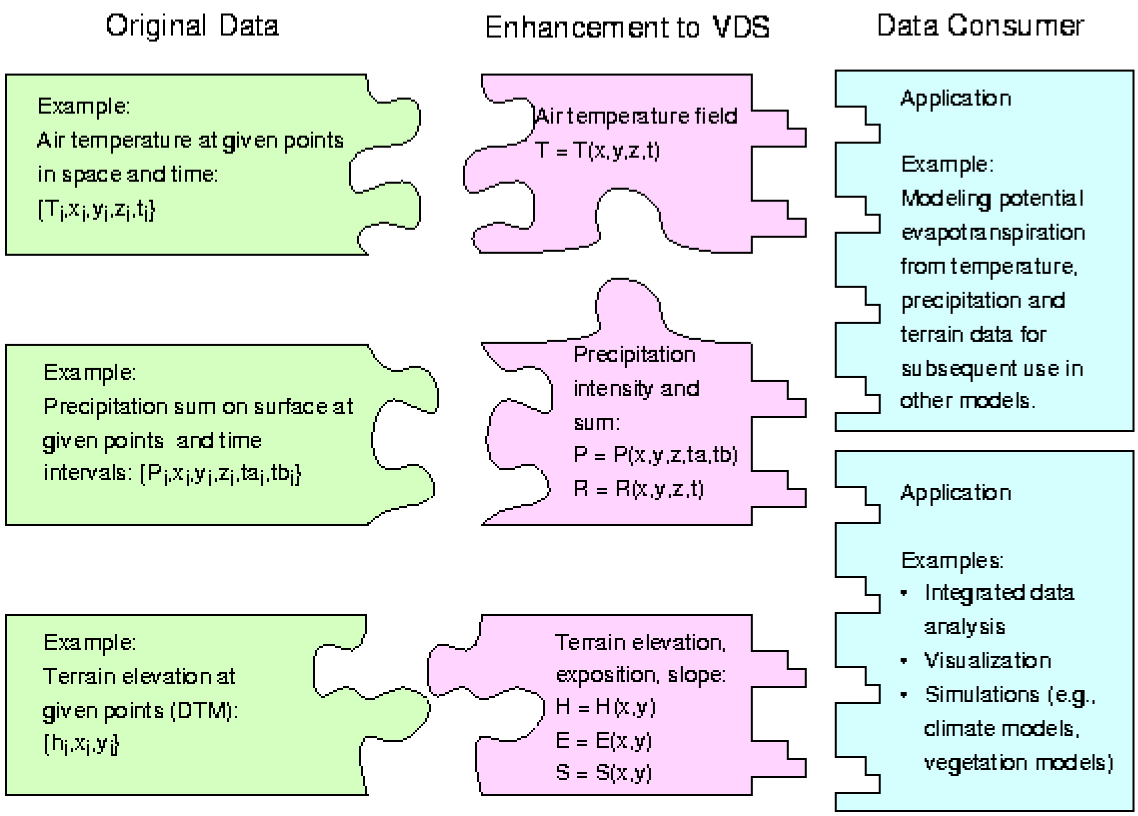
- The task might be very time-consuming and cumbersome.
- The transformation typically needs considerable process-based knowledge and mathematical and statistical skills. This is particularly true for appropriately selecting and parameterizing the model (Bucher & Vckovski, 1995).
- The model selection and construction is often restricted by the tools and methods available.
- Each transformation potentially amplifies the initial (and introduces new) uncertainties due to various reasons:

- A model is always a simplification of reality and is therefore always an *incomplete* image.
- Model selection and parameterization is subjective to some degree.
- Models and transformation methods are getting more and more sophisticated and their usage prone to errors.

The concepts of Virtual Data Sets (VDS) and Open Geodata Interoperability Specification (OGIS) discussed in the following section try to minimize the necessitity and drawbacks of subsequent transformations.

Virtual Data Sets and OGIS

The basic idea of VDS (Stephan *et. al.*, 1993, Vckovski, 1995) is to enhance a field representation with a set of methods which implement a model of the field. With these methods, a VDS is able to produce various *views* of the field depending on the data requirements of potential applications. VDS consists of *multiple, dynamically created* representations of a field. The methods are a *persistent* component of VDS whereas the values of a particular representation are *virtual* in the sense that they are derived upon an application's request. VDS encapsulate the field's *behaviour* and the (original) field representation and therefore represent an *object* in the sense of object-oriented design (Booch, 1991). With its ability to adapt to various data requirements (multiple virtual representations) a VDS can serve for many applications without need for conversions and transformations. Although these applications are required to access their views or representations through the interface offered by the VDS.



The OGIS Data Model (OGM) (Buehler, 1994) is an approach to define *and specify* a set of basic data types and their aggregates as building blocks for interoperable geoprocessing. The basic approach taken is to define *interfaces* to these data types (or classes). The OGIS approach is based on similar ideas but goes one step beyond VDS by providing a comprehensive *specification*. The specification consists basically of interface definitions written in CORBA's IDL (Interface Definition Language) and a set of assertions to restrict and define the semantics of implementations. Other approaches for the specification of OGM based on functional languages are discussed in (Frank & Kuhn, 1995).

The OGIS and VDS approaches envisage GIS technology based on (possibly distributed) interoperable objects where the classical gap between data structures and algorithms is bridged. Data sets are exchangeable components in the same way as other sets of services in a modular system. In combination with software technologies such as CORBA (CORBA, 1992), OLE, OSF/DCE or Java this leads to a true distributed geoprocessing environment.

Distributed processing has in the domain of environmental data management and analysis particular benefits as both data and specific processing methods are frequently exchanged and used within various organisations. More specifically, data users are often not in the same organisation than the data producer. The concepts of VDS and OGM allow data producers to create 'smart' data sets and to distribute them including the encapsulated methods. It can be expected that the quality of the dynamically created field representations is in general improved if data producers provide the corresponding methods. Data producers usually have most knowledge needed for the selection and parameterization of the methods.

The following section shows how Java can be used as a base for distributed geoprocessing and for the definition and implementation of VDS.

Java as an Environment for Interoperable Geoprocessing

VDS and OGM provide a conceptual framework for a component-based and true distributed geoprocessing. However, the *implementation* of distributed object systems is not a simple task. CORBA, OSF/DEC and OLE have become the major carriers in the software industry for such projects. Recently, Sun Microsystems presented the new programming language Java to the public. Java has a wide variety of promising features, some of them giving a powerful support for the implementation of distributed object systems.

Java is an object-oriented programming language and is syntactically based on C++. Many of the redundant and annoying features of C++ were removed based on the experience gained within the software industry in the past years (e.g., Java has no pointers). Java's popularity is particularly due to its usage within the World-Wide Web, as programming language for the development of so-called *applets*. The features of Java enabling its application within the World-Wide Web are the same which make Java a promising environment of distributed object systems. These features are:

- Java programs are compiled into a *portable byte code*. The byte code is interpreted on the target machine. Enabling the Java environment on a new platfrom consists of porting of the Java virtual machine (VM) (the byte code interpreter) to the target platform.
- *Late and dynamic linking* allows Java objects to be linked into a system at runtime. As an example, Java objects can be fetched over a network and then be linked to an application. Other approaches, such as CORBA, support similar concepts, although client-side execution fails due to the lack of portability.
- Java introduces the concept of *interfaces* (such as in *Objective C*). Interfaces are basically class declarations without any method implementation. They are *contracts* between an interface provider and an interface consumer on the structure of the interfaces. This complies with OGIS approach to specify their OGM with CORBA's IDL. A Java Class can implement (or inherit) many interfaces. There are no problems with ambiguity and multiple-inheritance such as in C++.
- Java provides a *verifyed and secure* environment for the exchange of small pieces of code and data across local and global networks. Mechanisms for the verification against tampering with object code are an integral component of a Java VM.
- Java's *packaging* allows a flexible naming scheme for components from various companies. A package defines a private *name-space* and therefore allows vendors to develop their Java-based components without interfering with other components.

A Java-based open and interoperable GIS-environment would consist of a set of Java classes containing both the (virtual) data sets and other geoprocessing services such as display, analysis and database management. For continuous fields, the corresponding Java classes contain both the (original) field representation as static data member as well as the transformation methods (e.g., for local interpolation, or calculation of derivatives) as class methods. Each VDS describing a field implements a fixed Java interface. Any application using the data needs only to have the field interface declaration to be able to use the data set. All low level details, e.g., on how the physical storage is organized, is hidden from the application.

A Java Interface for Continuous Fields

This example shall illustrate some of the concepts discussed above. It consists of three parts:

- Generic declaration of various types of field-*values*
- Generic declaration of field interfaces
- An example of a custom dataset based on the generic classes

The code presented here is by no means complete. Implementation bodies of most methods were left out for the sake of brevity. The class hierarchy was also reduced to a minimum to clarify the basic ideas.

The first part is a generic delcaration of the values of a field. Since all values are measurements, uncertainty and error information needs to be included. This is done by an interface called `ValueType`. It declares methods to retrieve some information on a "uncertain" value such as its mean, variance, etc. A more realistic example would be based on a set of different interfaces for various types of uncertain information, e.g., random variables, fuzzy sets, intervals etc. (see also Vckovski, 1995). `NormalValue` implements `ValueType` and declares a gaussian-distributed (random) variable.

```
public interface ValueType {

  float Mean();                 // Mean value of a measurement

  float Variance();             // Variance of a measurement

  float Min();                  // Minimum value

  float Max();                  // Maximum value

  float Median();               // Median value

  float Realize();              // Realization for stochastic simulations
                                // (if applicable)

}

// A normal (Gaussian-distributed) value
public  class NormalValue implements ValueType {

  float m,sigma;                // state: the 2 parameters of a gaussian
                                // distribution

  // constructor
  NormalValue(float amu, float asigma) {
    mu = amu;
    sigma = asigma;
  }

  float Mean() {
    return(mu);
  }

  float Variance() {
    return(sigma);
```

```
  }

  float Min() {
    return(-Inf);
  }

  float Max() {
    return(+Inf);
  }

  float Median() {
    return(mu);
  }

  float Realize() {
    return(Util.NormalRandomValue(mu,sigma));
  }
}
```

The follwing declaration defines a generic reference to a location on the earth's surface. It is in so far simplified as the ellipsoid used and other details are not referenced. A more comprehensive apporach would certainly also use additional represenation schemes.

```
// General geo-reference. Here simplified (e.g., 2-dimensional,
// no ellipsoid etc)
interface Georeference {

  float Longitude();

  float Latitude();

  void Set(float along, float alat);

  void Set(Georeference point);

}
```

These interfaces declare a generic scalar- and vector-valued field. The difference between scalar- and vector-field is made for the simplicity of using them. A scalar field is identical to a vector field with a value-domain dimension of 1, but it is simpler to use if it is declared as scalar value since no subscripts (array indices) need to be given.

```
// scalar field (= vector field with dim(Value-domain)=1)
interface ScalarField {

  ValueType Value(Georeference point);

  // in addition to this: functions to retrieve a description of the
  // support of the field
}

// vector field
interface VectorField {

  ValueType[] Value(Georeference point);

  // in addition to this: functions to retrieve a description of the
  // support of the field
}
```

The class `ScalarFieldOnGrid` is an example of an implementation of `ScalarField` for the case when the underlying representation is a regular grid. This class is *abstract*, i.e., it needs to be subclassed by an actual implementation.

```
public class ScalarFieldOnGrid implements ScalarField {
```

```
  Georeference origin;          // "lower-left" corner of grid (pt at 0,0)
  Georeference row1st;          //  pt at (0,1)
  Georeference column1st;       //  pt at (1,0)
  int rows,cols;

  abstract ValueType Value(Georeference point);

  // and here there would be a series of helper functions for the
  // management of data on regular grids, e.g., interpolation functions
  // etc.
}
```

The last part is a sample implementation of a Virtual Data Set describing the surface air temperature over a specifc area. It contains first a class defining the spatial references of this data set in a custom coordinate system (class `MyPoint`). The second class is the actual wrapper for the temperature data (class `Temperature`). Note that the data source can be an external data file, a database query (or view) or initialized static variables within the class. In that sense a Virtual Data Set can also be seen as *middleware* providing standardized access to environmental data.

```
// A custom georeference for this data set
public class MyPoint implements Georeference {

  float x,y;

  // construct via 'my' coordinates
  MyPoint(float ax, float ay) {
    x=ax;
    y= ay;
  }

  // construct via long/lat
  MyPoint(Georeference point) {
    Set(point);
  }

  float Longitude() {
    // de-project x,y
  }

  float Latitude() {
    // de-project x,y
  }

  void Set(float along, float alat) {
    // project along,alat to x and y
  }

  void Set(Georeference point) {
    Set(point.Longitude(),point.Latitude());
  }
}

// My temperature data set
public class Temperature extends ScalarFieldOnGrid {
  String datasourde = "http://myhost/mypath/tempdata";
  // or for example String datasource = "sql://mydbmshost/select tm,tstd,x,y
from temp;";
  // or static NormalValue[20][20] values  = { .... };
  NormalValue[][] values;
  Temperature() {
    origin = new MyPoint(42,42);
    row1st = new MyPoint(43,42);
    column1st = new MyPoint(42,44);
    // get data from 'datasource', fill up rows, cols, values[][];
  }
```

```
  NormalValue Value(Georeference point) {
    MyPoint pt = new MyPoint(point);
    // and now find out where pt is in the grid is and
    // perform interpolation, or fetch additional data if necessary
    // here is where the real work is done
    return (new NormalValue( /* */ ));

  }
}
```

The usage of such a VDS is sketched below. This can be either from within a Java-enabled generic GIS or any other tool for scientific computing.

```
  import ANETDataset.temperature;
// ...
  Temperature T = new Temperature();
// ...
  float avalue  = T.Value(8.5,47.4).Mean(); // this is Zurich ..
```

Conclusion

The objective of this paper is twofold. On the one hand, the discussion of some properties of environmental data has shown that modern technologies for object-oriented and component-based systems as they are adopted now in the GIS industry can be of great use in the management and analysis of environmental data. Particularly the representation of continuous fields can be significantly enhanced by embedding methods that model the behaviour of the natural phenomenon described by the sample data. On the other hand, this paper promotes Java as a carrier for distributed object systems in general and Virtual Data Sets in particular. Java's strength allows the implementation of interoperable and distributed geoprocessing in real and production-quality environments. However, the current lack of experience with Java in the GIS industry, and Java's future between market forces question its maturity for industry-strength software development.

Future work will focus on implementing the VDS concept as a subset of OGIS's OGM in Java. The experience gained hereby will help in the refinement of specifications of interoperable geoprocessing environments such as OGIS.

References

Albrecht, Jochen. 1995.
Semantic Net of Universal Elementary {GIS} Functions. *Pages 235-244 of: Proceedings of the AUTOCARTO 12 Conference*.

Booch, Gary. 1991.
Object Oriented Design with Applications. The Benjamin / Cummings Publishing Company, Inc.

Bucher, Felix, & Vckovski, Andrej. 1995.
Improving the Selection of Appropriate Spatial Interpolation Methods. *Pages 351-364 of:* Frank, Andrew U. & Kuhn, Werner (eds.) *Spatial Information Theory: A Theoretical Basis for GIS*. Lecture Notes in Computer Science 988. Berlin: Springer Verlag.

Bucher, Felix, Stephan, Eva-Maria, & Vckovski, Andrej. 1994.
Integrated Analysis and Standardization in GIS. *In: Proceedings of the EGIS'94 Conference*.
Buehler, K. A. 1994.
The Open Geodata Interoperability Specification: Draft Base Document. Tech. rept. OGIS Project Document 94-025. OGIS, Ltd.
CORBA. 1992.
The Common Object Request Broker: Architecture and Specification. Object Management Group. OMG Document Number 91.12.1.
Frank, Andrew. U., & Kuhn, Werner. 1995.
Specifying Open GIS with Functional Languages. *Pages 184-195 of:* Egenhofer, Max, & Herring, John R. (eds), *Advances in Spatial Databases*. Lecture Notes in Computer Science. Berlin: Springer Verlag.
Stephan, Eva-Maria, Vckovski, Andrej, & Bucher, Felix. 1993.
Virtual Data Set: An Approach for the Integration of Incompatible Data. *Pages 93-102 of: Proceedings of the AUTOCARTO 11 Conference*.
Vckovski, Andrej. 1995.
Representation of Continuous Fields. *Pages 127-136 of: Proceedings of the AUTOCARTO 12 Conference*.

WWW References

About Java
http://java.sun.com/about.html
OSF Distributed Computing Environment Home Page
http://www.osf.org/dce/index.html
Microsoft OLE Strategic overview
http://www.microsoft.com/devonly/strategy/ole/ole.htm

Karen K. Kemp

Easing environmental models into GIS

Part of the difficulty in integrating environmental models and GIS is a frequent mismatch between the digital data models used in each. Similarly, there is often a conceptual mismatch between the analog data models used by environmental scientists when t hey collect real world data and the digital data models offered by GIS. This paper argues that it is necessary to develop a clear understanding of the concept and role of the data model, and to recognize that there is a broad spectrum of different kinds of data models lying between the real world and the binary representations of it, including analog, conceptual and physical variations. Each transformation along this spectrum either removes information about the phenomena being represented, or adds inf ormation to the representation. While the database literatures gives us much background about the part of the spectrum comprising conceptual to physical data models, little research has been done on analog to logical transformations, an area of critical importance if GIS is to become a successful tool for environmental scientists.

Introduction

While the mathematical modeling of physical processes has been a focus of environmental scientists for at least a half century, implementation of these models has lagged behind their conception. Solutions have, of course, been found. For example, me thods for discretizing continuous equations are now well developed and finite differencing forms an integral part of many environmental modeling toolkits. The increasingly sophisticated tools for spatial modeling and analysis provided by today's GIS are n ow leading to a new revolution in environmental modeling, one which encourages scientists to incorporate spatial processes and relationships in their models. However, the driving force for the design of most widely used GIS packages has not been environm ental science. As a result, translation of the unique spatial concepts and models which have evolved independent of GIS in the various environmental sciences is not always obvious or without misapplication.

The relationship between the real world and how it is represented in GIS has been the subject of a number of important papers at these conferences and elsewhere (cf. Burrough and Frank 1995, Couclelis 1992, Csillag 1996, Goodchild 1992 and 1993, Kemp, 199 6, Nyerges 1991 and Peuquet 1990). In 1993, this author concentrated on how physical fields, which are a fundamental concept in much of environmental science, are discretized for representation within the digital computer (Kemp 1993). We now need to att end to the full spectrum of data models used by scientists and to the relationships between these and their digital representations. This becomes increasingly important as we move toward interoperability which by definition implies that there is a finite number of generic data structures appropriate for all realities represented. We need to understand how reality relates to such generic structures and what is missing in terms of models, metadata and encapsulated procedures.

This paper reviews definitions of data models and data modeling, discusses their significance in integrating GIS and environmental models, summarizes several efforts focusing on the data model issue and suggests needed research directions.

The many definitions of "data model"

The term "data model" was coined in computer science. The original definition may be attributed to Date who defined it as "a set of defined entities and the relationships between them" (Date 1975). There is little ambiguity in this definition and those working in the field of database management clearly use the term with confidence that the meaning is broadly understood. However, within this single field, it has been necessary to identify several classes of data models as one moves from reality to the digital world (Figure 1).

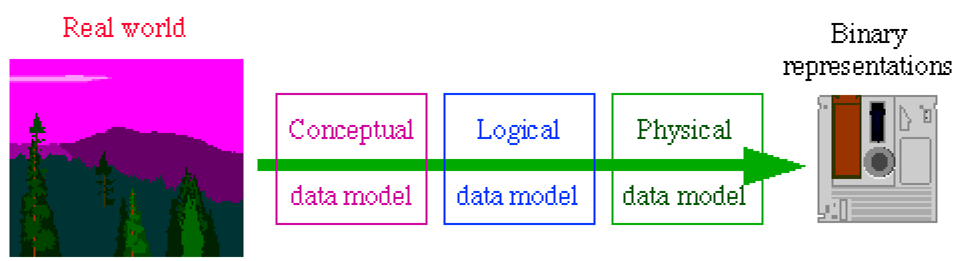


Figure 1 - Data models from database systems theory

Conceptual data models refer to entities in an enterprise and their relationships. Relational, network and hierachical data models are of the *logical* data model class. *Physical* data models refer to the digital structures used to or ganize and store the data within the computer.

However, although GIS textbooks frequently make reference to these standard DBMS terms, the term data model in GIS is used in a number of different ways and, as a result, confusion results. For example, data model is commonly used in GIS in the following contexts:

- vector and raster data models
- field and object data models
- representations for fields - pointgrids, contours, TIN data models
- even data structures such as GBF/DIME and chaincodes have been called data models (Peuquet 1990)

While all of these definitions are clearly related and similar, they are by no means synonymous. Each one addresses a similar issue, but from a different perspective. Without common agreement on what we mean, in GIS, when we say "data model", we cannot truly understand the fundamental importance of the concept.

Why do we need a better understanding of what we mean by "data model"?

Evolution in the understanding of these important issues is reflected in the hot topics arising from the three Environmental Modeling with GIS conferences hosted by the NCGIA in 1991 (Goodchild et al 1993), 1993 (Goodchild et al 1996) and 1996 (these proc eedings). Attention at the first conference was focused on the sophisticated uses being made of GIS in various environmental modeling disciplines. The second conference highlighted papers dealing with integration of data, individual GI systems and compu ter models of the environment. The third conference stressed interoperability--thus a concern with the integration of specific computing environments has been overtaken by consideration of the development of overarching theory and implementations through which everyone and every system is able to communicate with each other. The need for a common language is clear.

But why are the issues raised in discussions of data models important? While, as suggested above, there are many perspectives to the issue, all of the papers addressing the topic of data models seek to provide some further understanding of how we represe nt the world in a computer and/or database. Indeed, much progress is already being made in this direction. In (Kemp 1993), we outlined the need for a formalization of the relationship between the concept of a real world "field" and its representation in one of 6 different spatial data models (here referring to grids, polygons, TINs, contour lines, pointgrids and irregular points). Each of these representations imposes different interpretations of the continuous nature of the field and implies different techniques for interpolating values at points between those few for which data is stored in the digital database. We argued that it is necessary to retain certain information about the relationship between the reality being represented and the model use d to store it. Some of this information is implicit in the data model chosen, some of it must be explicitly stated (eg. as encapsulated operations).

In a guest editorial in *International Journal of Geographical Information Systems*, Burrough and Frank have mused upon the importance of understanding "the philosophical and experiential foundations of human perception of geographical phenomena and their abstraction and coding in geographical information systems" (Burrough and Frank 1995, p. 101). They consider how "geographic data models" reflect how people view the world. While they do not specifically define what they mean by their data model t erm, the discussion incorporates a consideration of spatial data paradigms and the variable aspects of representations affected by differences arising from different user communities and cultures. They conclude that:

> the question arises of how one can sensibly integrate different kinds of spatial data if each has been observed, recorded, modelled and stored according to its own particular set of paradigms.... The main conclusion must be that methods of ha ndling spatial information must be linked to the paradigms of the users' disciplines and that inter-disciplinary research to determine more accomodating paradigms than the object-field models is essential. (Burrough and Frank 1995, p. 114)

What does a data model do?

If we are to uncover what we are trying to get at by using the term data model, it is useful to express what a data model is intended to do. Writing in the database management literature, Brodie suggests that:

> a clear goal for a data model is that it be expressive. Using the data model, one should be able to represent any static or dynamic property of interest to the desired degree of precision in order to capture the intended meaning (Brodie et al . 1984, p. 41).

Similarly, Goodchild and others have recently written that "in essence, a data model captures the choices made by scientists and others in creating digital representations of phenomena, and thus constrains later analysis, modeling and interpretation" (Goo dchild et al. 1995, p. 10).

Data modeling is the process by which entities in the real world are discretized. While sampling the real world requires abstraction so that the natural complexity can be reduced to simple data, data models allow the addition of information to raw data. For example, the TIN model allows a network of points to be joined in such a way that sloped surfaces are represented. Thus, complexity is returned to simple data.

Therefore we suggest that the term data model be understood within the GIS context to exist across the full spectrum between the real world and its binary representation. Thus data models may include reference to any of the following:

- data structures - points, polylines and polygons, raster and vector
- representations of fields as pointgrids, polygons, contours, TINs, cellgrids, irregular points
- a link and node network representing a hydrological system
- a transportation network used as an addressing system
- abstract discrete models identified simply as fields or networks
- and importantly, analog data models such as those used during field sampling and data collection.

Moving from the real world through various data models to model output requires transformations in both information structure and information content. These transformations from the real world to binary representations of it include:

- abstraction, generalization and selection of relevant concepts, processes and relationships in the real world
- conceptual modeling of the relationships between abstract entities
- mathematical modeling of the relationships between defined entities
- physical sampling of the real world
- storage of data in computers - may or may not include the necessity to model space
- transformation of data between different representations (models).

Transformations take place as real world data is collected, recorded, manipulated and eventually stored in digital databases. Along this transformational path, the people who are manipulating the data transition from environmental scientist to computer an alyst to "naive" user. How much real world knowledge can be passed from each of these individuals to the next through the data itself? As the data model and, possibly, associated metadata and functionality or procedures are the media, it is critical tha t these incorporate all the relevant information.

As described earlier, the transformation from the real world through data models to binary representations involves some loss and, sometimes, recovery of information. There are a number of dimensions in the information which may be captured or lost in th e data modeling process (after Burrough and Frank 1995):

- dimensionality - fields or objects, 2D, 3D
- scale - single- or multi-scaled
- representation - generalized to detailed
- exactness - can phenomena be exactly described?
- logic - single value (Boolean) or mulitvalued (fuzzy)
- static or dynamic
- measurement scale - nominal, ordinal, interval or ratio
- enumeration - complete or sampled
- deterministic or stochastic

Analog data models versus digital data models

Analog data models define fundamental primitives which conceptually discretize the infinite complexity of reality. However, unlike digital data models, analog data models can be continuous and they may or may not include the same primitives used in the d ata models to represent the same phenomena digitally. Data collected in analog field data models include:

- points on lines
    - geology - transects recording significant rock type transitions
    - hydology - stream cross-sections
    - oceanography - transects taken from ships
- areas
    - ecology - sample plots
    - soils - homogeneous areas defined from air photos
- points
    - geology - boreholes
    - hydrology - well logs (which provide a continuous measurement of a changing surface)
    - soils - soil pits

The point is that while the data may be discrete, the analog model relates these discrete values to continuous mental models of the real world.

Interoperability through data models

Data models provide entities and relationships at various levels of definition and discreteness. Interoperability requires the identification of what information is provided by the specific data models used in each computing component, a method for trans ferring that information with as little noise as possible plus, ideally, some measure of the amount of information lost in any processing stage. Interoperability through generic data models are described by papers presented in this session. Vckovski and Bucher lay out a specification through which information about data and data models can be included in the Open Geodata Interoperability Specification (OGIS). Albrecht supports interoperability by identifying a generic set of functions which operate on standardized data models.

In another major project addressing related issues, Smith and others are attempting to implement a computational modeling system (CMS) "which is intended to provide scientific investigators with a unified computational environment and easy access to a bro ad range of modeling tools" (Smith et al. 1995, p. 127). It is an impressive effort and sets in place many fundamental programming concepts necessary for interoperating modeling environments. In particular, it outlines a process by which symbolic repres entations of phenomena or concepts can be constructed from fundamental primitives and provides a means for relating these different concepts through transformations. The system allows the existence of both abstract representations, which may simply ident ify instances of a particular concept by name, and mulitiple concrete representations of these which are their various digital representations. Existing in an object oriented programming environment, concrete representations are built up inductively from primitives and previously defined super- or parent-representations.

While this CMS promises to provide an extremely powerful and flexible tool for environmental scientists conducting a modeling project, it does not adequately address the conceptual end of the data modeling process. No support is provided to assist the en vironmental scientist in formulating the abstract representations of the phenomena and concepts being studied. The system requires representations to be built up from primitives, but how does one go about identifying the concepts and their transformation s and relating them to appropriate primitive? What is still missing is a mechanism for ensuring the appropriate information is passed from the real world to the data models.

Data models and GIS research

Given all of the above, it should be clear that data models provide fundamental areas of investigation in many GIS research environments. For example, within the NCGIA's research agenda, we can find the following very different considerations of the data model problem:

- Initiative 15: Multiple roles for GIS in US Global Change Research
  - In order for digital data and geographic information tools to be generally useful for the global change community, there is a need to understand the spectrum of different data models used by these scientists.
- Initiative 17: Collaborative Spatial Decision Making
  - Data models of different decision makers must be intergrated in order to compare different positions.
- Initiative 20: Naive Geography
  - Naive Geography is defined as the field of study that is concerned with formal models of the common-sense world.

More specifically, the research agenda on data models from the first I-15 Specialist Meeting includes:

- characterization of existing different models
- reconciliation of different models
- methods to measure the representational efficiency of a data model
- metrics for measuring user satisfaction with a data model
- a common underlying idea of time and space on which data models can be based
- support for a representation of change in a data model
- development of a language for data model description
- documentation of models from each domian
- abstraction and identification of shared elements of data models
- construction of translators between models at the conceptual level

Conclusion

Much progress has been made in recognizing and structuring the information content of digital data models used in GIS. However, much more effort is needed in understanding analog spatial data models (i.e. the models used by environmental scientists) and their relationship to existing and future digital spatial data models.

Acknowledgements

Research at the NCGIA is supported by a grant from the National Science Foundation (SBR 88-10917).

References

Brodie, M. L., J. Mylopoulos, J. W. Schmidt. (1984). *On conceptual modelling : perspectives from artificial intelligence, databases, and programming languages*. New York, Springer-Verlag.

Burrough, P. A. and A. U. Frank (1995). Concepts and paradigms in spatial information: are current geographical information systems truly generic? *International Journal of Geographical Information Systems* 9(2): 101-116.

Couclelis, H. (1992). People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS. *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*. A. U. Frank, I. Campari and U. Formentini, Springer-Verlag. 639: 65-77.

Date, C. J. (1975). *An Introduction to Database Systems*. Reading, MA, Addison-Wesley.

Csillag, F. (1996). Variations on hierarchies: Toward linking and integrating structures. *GIS and Environmental Modeling: Progress and Research Issues*. M. F. Goodchild, L. T. Stayaert, B. O. Parks, C. Johnston, D. Maidment, M. Crane, S. Glendinni ng, eds. Fort Collins, CO, GIS World Books: 433-437.

Goodchild, M. F. (1992). Geographical data modeling. *Computers and Geosciences* 18(4): 401-408.

Goodchild, M. F. (1993). Data models and data quality: Problems and prospects. *Environmental Modeling with GIS*. M. F. Goodchild, B. O. Parks and L. T. Steyaert. New York, Oxford University Press: 94-103.

Goodchild, M. F., B. O. Parks and L. T. Steyaert. (1993). *Environmental Modeling with GIS*. New York, Oxford University Press.

Goodchild, M. F., J. E. Estes, K. Beard, T. Foresman, J. Robinson (1995). *Research Initiative 15: Multiple Roles for GIS in US Global Change Research. Report of the First Specialist Meeting*. Santa Barbara, CA, National Center for Geographic Infor mation and Analysis, University of California.

M. F. Goodchild, L. T. Stayaert, B. O. Parks, C. Johnston, D. Maidment, M. Crane, S. Glendinning, eds. (1996) *GIS and Environmental Modeling: Progress and Research Issues*. Fort Collins, CO, GIS World Books.

Kemp, K. K. (1993). *Environmental Modeling with GIS: A strategy for dealing with spatial continuity*, National Center for Geographic Information and Analysis, Department of Geography, University of California, Santa Barbara.

Kemp, K. K. (1996). Managing spatial continuity for integrating environmental models with GIS. *GIS and Environmental Modeling: Progress and Research Issues*. M. F. Goodchild, L. T. Stayaert, B. O. Parks, C. Johnston, D. Maidment, M. Crane, S. Glend inning, eds. Fort Collins, CO, GIS World Books: 339-343.

Nyerges, T. L. (1991). Geographic information abstractions: conceptual clarity for geographic modeling. *Environment and Planning A* 23: 1483-1499.

Peuquet, D. J. (1990). A conceptual framework and comparison of spatial data models. *Introductory Readings in Geographic Information Systems*. D. J. Peuquet and D. F. Marble. London and Bristol, PA, Taylor & Francis: 250-285.

Smith, T. R., J. Su, A. El Abbadi, D. Agrawal, G. Alonso, A. Saran (1995). Computation Modeling Systems. *Information Systems* 20(2): 127-153.

Karen K. Kemp
Assistant Director, National Center for Geographic Information and Analysis
University of California
Santa Barbara, California 93106-4060
Telephone: (805) 893-7094
FAX: (805) 893-8617

Kenn Gardels

The Open GIS Approach to Distributed Geodata and Geoprocessing

Open geographic information systems represent an evolution from traditional GIS solutions, in which proprietary data models and monolithic software functions are made interoperable and extensible. Applications which adhere to the objectives of Open GIS are more able to access and use various types of distributed data, and to utilize multiple geoprocessing tools and services. A formal specification, the Open Geodata Interoperability Specification (OGIS), is currently under development and will define the types and methods necessary to build interoperable systems.

Major areas of research and development within OGIS include defining a geodata type hierarchy that comprises simple and complex features, maps and coverages, images, and field datasets (and specifying the relevant interfaces); developing a consistent approach to metadata to support data collections and browsing; enabling the concept of ``information communities'' which share common definitions, information semantics, and data dictionaries/thesauri; and defining the basic data structures required for implementation.

In the context of environmental modeling, OGIS will provide a significant gain in interoperability between conventional geographic data stores and modeling system, through the use of a common language for sharing geodata and standardized definitions of interfaces to functions which operate on geographic information. While initial work is focussing on traditional geoprocessing services, such as spatial selection, thematic overlay, measurement, and distance analyses, other services which access geographic information, such as hydrodynamic models, seismic prediction, and allocation functions, will also be able to directly access geodata stores (as well as other geoprocessing functions). In support of this, a major new research effort must be undertaken with respect to temporal data modeling and the effective integration of spatial and temporal data types and methods.

What is Open GIS?

As costs of computer hardware and software for geographic information systems (GIS) decline, resource managers are paying increasing attention to maximizing the value of environmental data. Much work to date has been done in the areas of standardized data development, distribution of digital information, and format translation. Less effort has been directed toward the exchange of information between inventory-based GIS applications and analytical tools such as statistical analysis, process modeling, and pattern recognition. Future success of GIS as a technology and as a paradigm of spatial understanding will depend on the seamless integration of diverse methods into a comprehensive system for scientific investigation and environmental planning.

The open systems model is an approach to software engineering and system design that enables and encourages sharing of data, resources, tools, and so forth between different users or applications. When applied to the domain of geographic information systems, the intent is to move away from the current paradigm in which specific GIS applications and capabilities are tightly coupled to their internal data models and structures. Open GIS facilitates exchange of information not only between individual GISs but also to other systems, such as statistical analysis, image processing, document management, or visualization.

The basic limitation of geographic information management being addressed by open GIS is the proliferation of different data types, geographic information systems, and applications. Such diversity is characteristic of a rapidly developing field with a seemingly unlimited set of users and data requirements, but it works against meaningful sharing of data resources. Users who wish to gain access to geodata developed by others are generally faced with a complex data conversion task. Adoption of transfer formats helps this situation, but does not solve it entirely due the the complexities associated with merging vastly different data models and schemas. Toolbox approaches to GIS seek to address this problem by providing a wide range of geoprocessing tools operating on a private data model. Adopting a private data model, though, necessarily limits users to the tools available within the toolbox; they cannot freely utilize other types of software functionality, such as statistical analysis, spreadsheets, or numeric processing. Application developers must often accommodate fundamentally incompatible data models and geoprocessing tools by creating custom software for specific functions.

The fundamental requirements of an open GIS, that addresses these issues, are:

- interoperable application environment - a user workbench that is configurable to utilize the specific tools and data necessary to solve a problem
- shared data space - a generic data model supporting a variety of analytical and cartographic applications
- heterogeneous resource browser - a method for exploring and accessing the information and analytical resources available on a network

The Open GIS Consortium, a broad-based alliance of government agencies, research organizations, software developers, and systems integrators, is engaged in a multi-year effort to define open GIS and to develop a set of requirements, standards, and specifications which support it. The overall goal is to encourage software developers and integrators to adhere to these requirements, and through time create tools, databases, and communications systems that maximize the utility of systems and resources and take advantage of technological advances. As noted in the *Open GIS Guide*, the objective is technology that will enable an application developer to use any geodata and any geoprocessing function or process available on ``the net" within a single environment and a single work flow.

An Architecture for Open GIS

The three broad requirements for Open GIS - interoperable applications, shared data space, and heterogeneous resource browser - must all be linked into an overall system architecture. Although each may be seen as a distinct set of capabilities, they all must coexist in a common framework that defines how system components interact. Of course, these components are themselves complex and have multiple levels and modes of interaction with each other.

An open GIS architecture must provide robust methods for accessing multiple forms of data using multiple software environments. That is, any compliant GIS or other application that uses geodata must be able to access and use distributed information in any supported format. The Open GeoData Interoperability Specification (OGIS) project, underway since June 1993, is an attempt to design methods that provide an object oriented architectural framework for access to geodata, independent of the specific data structures and file formats used to model the data. From a user's point of view, OGIS allows access to geodata at remote locations, no matter the format. From an application developer's point of view, OGIS provides a set of network services to identify, interpret, and represent a dataset from a geodata server to a geoprocessing client.

Figure 1 provides an overview of the OGIS architecture, consisting of distributed computing platforms, data objects and stores, and applications and services. Although Figure 1 might suggest a layered architecture, this is only to diagram the operational relationships among the various components. In fact, every construct in this architecture - GIS database applications, geoprocessors, framework components -

is an object. Each has specific interface methods which allow it to communicate with other objects. What makes this approach unique is that specialized services are being developed which provide the functionality needed for interfaces to be meaningful. For example, access providers utilize the detailed protocols necessary to represent a private data format within the generic geodata model of the framework. In fact, Figure 1 could be redrawn with ``distributed object technology'' as a bus, to which are connected all other functional constructs in the form of objects providing and requesting services.

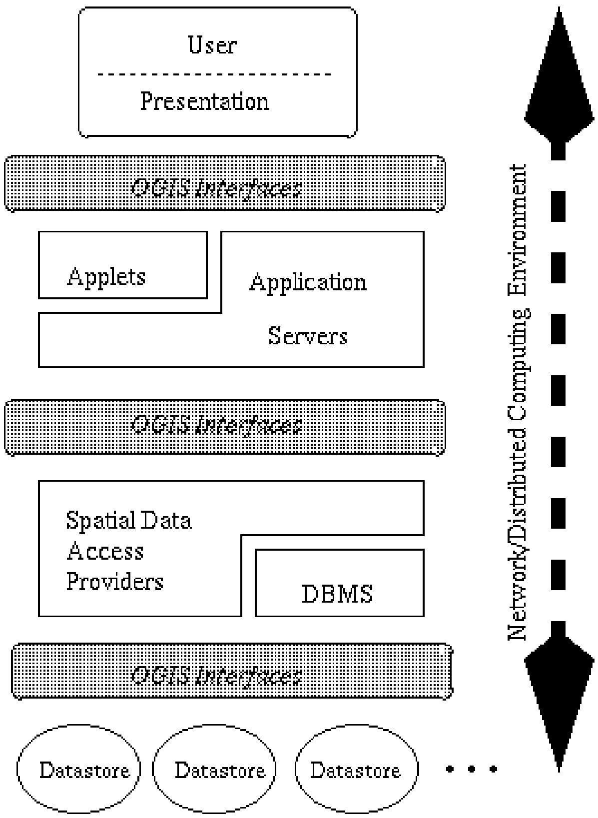


Figure 1. Interoperable Geoprocessing Model

It is important to note that OGIS is an operational model, not a data standard. The overall premise of OGIS is a specific set of software tools for dynamically translating geodata from various sources into a single, comprehensive object based data model which can then be accessed directly from applications using a basic toolkit or primitive operations. In the context of open geographic information systems, OGIS provides the framework upon which rests other broad categories of functionality, notably the geodata model which describes geographic information and the services model which which defines the interfaces to applications software and systems.

OGIS attempts to accomplish these operational objectives, first, using an object-oriented implementation, based on existing distributed object methods and architectures. This means that every component of the specification - be it a data format, an application, a geodata model, a converter, or a user schema - is an object and as such can be manipulated by a set of utilities designed for object management. Second, OGIS development will be phased both in a general architectural sense and in its specific prototype implementations. That is, methods will be established initially to provide simple exchanges of geodata, followed by operators on various formats, and finally by a incremental approach to defining and implementing geoprocessing services. Third, OGIS development will start with well-defined, open interchange and transfer formats as a basis for geodata objects, such as SDTS. As procedures for interpreting these formats within the OGIS model become well-defined, subsequent efforts will focus on standard distribution formats and (with outside system developers) private data formats.

A key characteristic of OGIS is determined by the plethora of existing data structures - not only must specific translations be robust and consistent, they must resolve inconsistencies between various data models. For example, OGIS must enable transparent access to both geo-relational and feature-based geodata models. This key is interoperability - the ability to access and translate data based on a process of discovery and dynamic interpretation when the salient factors cannot be known in advance.

The OGIS process is intended to unify the multiplicity of approaches to geographic information system technology in three key areas: geodata models, geoprocessing services, and defined information communities. Together, these efforts will bring traditional GIS practice, automated mapping, remote sensing, spatial and temporal analysis, and scientific modeling to a common geographic information framework.

Shared Data Model and Store

The OGIS Geodata Model (OGM) is at the heart of the open GIS concept, and represents a integral part of the OGIS framework. It provides a consistent logical view of geographic information, independent of the underlying data model or format. Because it is a comprehensive geodata representation, it allows the creation of a set of high-level functions or operations which can be used by applications for accessing different data sets.

Geographic information is collected and managed for numerous purposes, each of which has its own requirements for how data are most efficiently organized, what comprises features of interest, what degree of precision and accuracy is necessary, how information is analyzed and displayed, and so on. As a result, there are now many geodata representations, which are largely incompatible and which limit their utility for a community of users. For purposes of this discussion, we define the following levels of representation of geographic information:

- data model refers to the conceptual view of a set of information, for example map themes, discrete features and objects, observations, or numeric or algorithmic descriptions.
- data structure refers to the method used to encode geographic information, such as arc-node (topologically related vectors), CAD, raster, database records, or linked objects.
- data format refers to the specific protocol or procedure used to store and manage geodata, such as GRASS run-length encoding for rasters or ARC/INFO coverages for vectors.

In these definitions geographic information comprises data points or features in which every entity can be uniquely described in terms of its physical location in space (and possibly time). It is true that there are many other forms of information that relate generally to space, such as a journal article on a place, but these types are not included in this working definition of geographic information, except as described below.

The objective of the OGM, then, is to create a single comprehensive model which embraces the range of existing models and their associated structures. That is, the OGM must be able to describe any datum held in any format developed to the parameters associated with any data model. From an application, rather than data, perspective, the OGM framework must provide methods by which a user can query geographic information.

Initially, the OGM must address data management requirements for commonly used geoprocessing systems and the geodata associated with them. These include not only a variety of formats and structures, but fundamentally different models for representing geographic phenomena. Thus an OGM structure must be defined from the top down to at least provide a context for divergent data structures.

At the top, conceptual level, an OGIS geodata object comprises three major components:

- Spatial components composed of geometries, such as points, lines, polygons, grids, and spatiotemporal referencing, defined by projections, coordinate systems, and allowable transformations.
- Semantic components defining the meaning of an object's elements in terms of a ``real-world'' model, using a catalog or data dictionary.
- Metadata describing any additional information required to interpret an object correctly in the context of an information community.

OGIS recognizes that there are two fundamentally different ways in which we view the earth -- as entities and as phenomena. Entities are discreet, identifiable constructs or units that have relatively well-defined boundaries and unambiguous descriptions, such as buildings, water bodies, or measurement stations. Phenomena vary more or less continuously over space and their description is meaningful only at a particular point, for example topography, air quality, or soil type. This duality is captured in OGIS as two geographic types:

- Features - represent atomic, uni-valued geographic entities, whose location is a function of their definition, ie ``where is what'' representation.
- Coverages - describe complex geographic phenomena in terms of maps, images, and fields, and are a function of the spatial/temporal domain, ie ``what is where.''

[In the rigorous definition of the actual specification, coverages are a special sub-type of feature, to better manage common interfaces between feature collections and coverage information.]

OGIS objects are described by a set of class definitions and method interfaces (a class library in an object-oriented programming terms; a schema in OO-DBMS terms). Use of an object oriented database structure and programming approach is necessary to meet the objectives of the OGM. This is true for the following reasons:

- the OGM must provide a mechanism for storing and managing data in a wide variety of formats associated with multiple data models. This cannot be accomplished using traditional database stores due to their limitations of datatypes or using typical GIS datafiles due to the lack of integration between those datafiles.
- the OGM must allow random access to any selected data element based on a simple description so that information retrieved from private datasets is immediately available for use without further processing. This cannot be done using most data transfer formats (such as SDTS or DLG) since geographic features are not uniquely identified or completely described.
- associated with each type of geographic object are methods for selection, analysis, display, and combination with other objects. Object oriented systems provide this close coupling of datatypes and methods.
- it must be possible to provide persistent storage of modeled geodata for continued use without explicit intervention by the user. That is, when information is remodeled in response to a query, the resulting objects should remain in the store for later use. This cannot be done in conventional procedural programming without creating an independent file storage system, which negates the purpose of the OGM.
- the system must be fully extensible, so that new data types and methods can be accommodated without modification of the existing system.
- the OGM must manage not only geographic information as defined above, but also be able to link this with descriptive metadata, which may be explicitly derived from private data or inferred from other sources. Although conventional relational DBMSs provide ad hoc query construction, they presuppose a pre-defined schema for all possible data values and types; OO systems allow such relationships to be defined on an as-needed basis.

The OGIS Geodata Model may be viewed as a type hierarchy, as summarized in Figure 2. The generalized notions of dataset and feature may be decomposed into their constituent spatial, semantic, and metadata components. The geometries of spatial components reflect well-accepted approaches to geodata modeling as well as the dual nature of simple or complex features and sets versus coverages. The key aspect of spatial reference is that rather than being subsumed within a metadata tag about a dataset, the algorithmic definitions and transformations are closely coupled to the feature's geometry. Semantic components are defined by an attribution model, such as the georelational model typically used in current GISs, plus the collection of properties and their parameters associated with the feature. Metadata are properties pertaining to feature sets or collections (and to coverages when evaluated in this manner) that apply uniformly to all the individual elements within a set; applied recursively to complete data stores or complex systems metadata define features from the highest level of aggregation down to the atomic level of specific data elements.

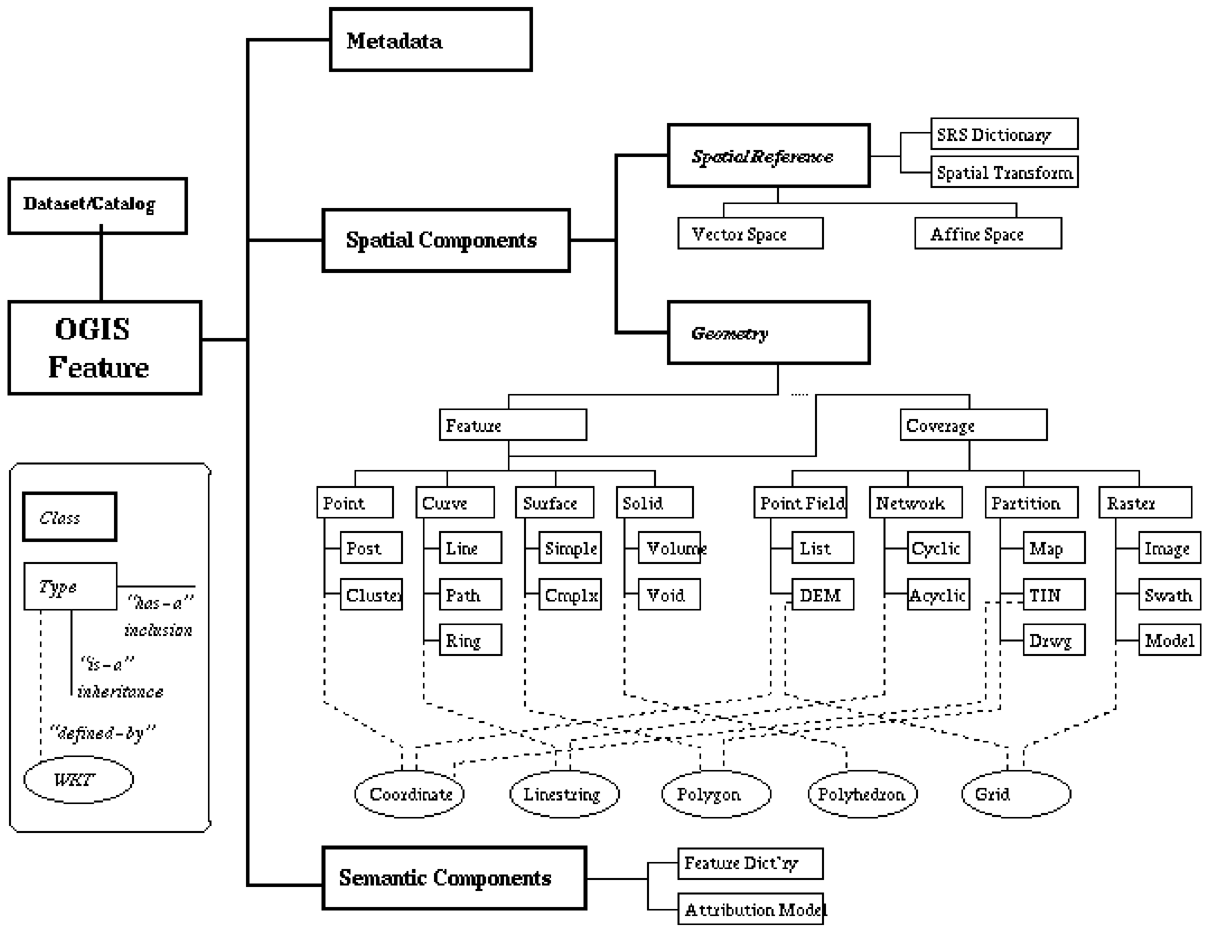


Figure 2. The OGIS Geodata Model

Services and Applications

Along with the OGIS GeoData Model, the Services Model comprises the operational, process-oriented aspect of OGIS. The services model consists of a fundamental set of functions which facilitate the creation and implementation of applications and information communities. Consistent with the object orientation of OGIS, these functions are realized as types -- software objects that provide specific functionality. The types specify the interfaces that applications use to access other applications and services.

Services in the OGIS architecture are specific instances of applications with the additional characteristic that they provide capabilities for client applications and services. For example, services enable access to a variety of private data formats, provide specific transformation or geoprocessing operations, and facilitate identification and location of resources available in a distributed environment.

Applications of the OGIS architecture are processes, running on some location on a network, that respond to a specific set of user requirements. For example, a hydrologic modeling application could obtain spatial information from multiple geographic data bases. Applications might also establish a mechanism for integrating existing tools and datasets into special-purpose, high-level decision support or modelling systems.

Catalog and Schema Services

The essential service in the framework is the ability of applications to request and utilize data in different formats and storage systems. This means that OGIS compliant applications can use data held in common transfer standard or distribution formats or in GIS databases. The geodata access capability provided by OGIS is independent of underlying data models and structures, since each supported private data store is transformed according to the definitions of the OGM; that is, each store is rendered to the application via the defined interfaces. The underlying principle of OGIS geodata access is to provide a uniform approach to obtain information that does not require users to know or understand the original organization of the data.

The mechanism for data access is a series of access catalogs that provide information on the content of the data within a private store, describe the OGM types that can be created from the private data, and supply a handle or pointer to the selected data, acknowledging the feature schemas associated with both the source and the target. Closely related to the underlying data structure and content is the extent to which those geodata can be transformed to the OGM. This is a function both of the implicit nature of the source geodata and of the specific capabilities of the associated data provider. An application must be reliably informed of exactly which OGM classes can be instantiated and what interfaces are available.

The process also includes information about the semantic content of the data objects, directly through a shared information community schema, or indirectly via a semantic translator service with pre-defined mappings. The exchange of information may include spatial and temporal metadata - information about the geographic dataset - as well as both the spatiotemporal reference and thematic meaning of geographic phenomena. Due to the fact that GIS databases and archives are created and maintained using an almost limitless range of schema designs, data dictionaries, content glossaries, and gazetteers, applications must be able to obtain meaningful information as to data content and semantics. This information is the basis for formulating queries and interpreting results. The geodata catalog service reports database content in a structure and at a level of detail appropriate to the applications' requirements.

Once an application understands the utility of the private data, it can negotiate with data providers to make data available (via OGIS interfaces) that have been exposed by the catalog service. This is essentially a read operation, where a previously discovered geodata item is rendered as an OGM object (or collection of objects). From the perspective of the private data store, this involves selection of a specific set of geodata. Selection is based on criteria related to the thematic and/or spatial content of the geodata, and can be very specific or highly generalized. [It is up to the query service (see below) to indicate what selection criteria can be used, and how they will be evaluated.] Spatial selection can be keyed to a general location, coincidence with other geographic phenomena, or locational/topological relationships. Similarly, thematic selection can relate to a general theme, specific information content, or aggregations and associations. It is up to the data provider to define the range and specificity of allowable queries, falling somewhere along the continuum suggested by the following scenarios:

- The only selection supported is defined in terms of the private data store's internal organization, i.e. map coverage, raster file or drawing, and file name or title.
- An arbitrary bounding box can be applied to limit the spatial extent of geodata, along with specification of the feature type (e.g., map theme).
- Feature definitions can be used to determine the particular entities to be selected, such as geometric representation, spatiotemporal relationships, thematic content, or any other attribute managed and explicitly stored in the source system.
- Other geodata, either in the source system or modeled in the OGM, can be used to determine the spatial extent of selected data, based on computation of regions of coincidence, proximity, and the like.

Transformation Services

Catalog and access services are able to make selected data available via specified OGIS interfaces, but are unable to actually change one data model to another. Applications may need to obtain geodata in a completely different model or structure than that of the source store. For example, an image processing application may require that polygonal feature information be rasterized before it can be used. Although both models are supported by the OGM, the user is effecting a transformation process. In addition, geodata are frequently captured and stored in a different geographic projection, coordinate system, or geodetic reference system than needed for an application, especially where multiple map layers are involved. The OGIS transformation service is a program, available in a networked environment, that provides a function or set of functions that convert data from one OGM data model or geodetic framework to another (within the OGM). The transformation service can be invoked either by an access manager as part of the geodata access process, or directly from an application. Semantic transformations are required, as described above, to associate feature definitions in one schema or data store with those in another.

Query and Geoprocessing Services

Query services provide mechanisms by which applications' requests for data are evaluated and executed, and they manage the return of query results to the initiator. Query services are based on a standardized query language that allows applications to access spatiotemporal, semantic, and descriptive aspects of a feature or feature collection. The language must provide primitives for basic spatial operations such as intersect and clip, semantic operations such as selection by range or equivalence, and descriptive operations using keywords or parsing. Data providers, depending on their level of compliance with OGIS, are able to provide limited geoprocessing, such as feature selection and possibly overlay. More complex analytical operations, such as map algebra, buffering, classification and filtering, and similar GIS kinds of operations, are generally outside the purview of data stores. Although an application could perform these operations directly, distributed object technology provides an opportunity to utilize specialized services for needed functions. The role of geoprocessing services is to provide applications with high-level tools that are not necessarily available within the application. These network resources can function as a GIS toolkit which is rendered accessible to any requesting client process. They may be implemented as basic services, where their functionality is seen as fundamental to OGIS operation, or as vendor-supplied utilities which meet specialized requirements. In the latter case, the service could be provided by a conventional GIS application with a public interface for request and response, possibly in conjunction with a GIS database. For example, an application may request geodata objects comprising all parcels within 10 kilometers of designated water bodies; a combined GIS database access manager and geoprocessing server could provide both the necessary data and the buffering operation. Key to most geographic analyses and queries is visualization of results in the form of interactive display or hard-copy output. Applications may enlist a display service that can render drawings and images. A more intelligent display service could also address cartographic symbolization and generalization, projection and transformation, and three dimensional animation.

Information Communities

The OGIS concept of Information Communities address issues of semantic interoperability; that is, the ability of users to share meaningful information about the earth. It recognizes that different groups of users may have profoundly different world views, each with its own vocabulary and taxonomy, abstractions of earth features and phenomena, and accepted mechanisms for manipulation, communication, and analysis. Within a given information community, sharing of information is relatively straightforward, with only minimal requirements for communicating encoding protocols, dictionary structures, etc. Between two or more information communities, information sharing involves translators that convert or filter the semantic content of each community's store into a common representation or view.

The semantics associated with an information community (as listed in the OpenGIS Guide) include: metadata to describe the content of feature collections; feature class definitions; attribute definitions; valid feature and attribute relationships; data capture guidelines for features and attributes; symbol sets for feature representation; rule sets for feature portrayal or display; relationships and dependencies between features, attributes, metadata, etc.; methods; and behaviors. Furthermore, each distinct and unique geodata collection is owned by exactly one information community, created according to the semantic rules associated with that community. An information community is also defined by the availability of one or more catalogs, comprising metadata describing each of its feature collections; such catalogs may be organized as a hierarchy of increasingly generalized descriptions of sets of individual catalogs for specific feature collections.

If an information community's catalog is sufficiently robust, a user in another community should be able to use the catalog to determine the precise meaning of the collection's components and thereby utilize that information for its own purposes. The only requirement is that the source catalog be exposed to the world through simple avenues like the world wide web or more complex approaches involving object trader services in a distributed computing environment. Alternatively two information communities may agree to cooperate in developing a semantic translator service that provides a formal, explicit mapping between the feature types in one to those in the other. In effect, this means creating an alternate view of a collection that is consistent with the world view of a different user.

Issues of database federation and schema fusion are not unique to the geodata/geoprocessing realm. Therefore future efforts in the area of OGIS information communities will be guided by developments in the distributed processing industry generally. At this writing, the information community concept is evolving steadily, so the reader is referred to on-going discussions and proposals at the OGIS web site.

Environmental Modeling Using OGIS

OGIS is intended to support interoperability between multiple models of geographic information and a broad spectrum of analytical and visualization applications. To date, much of the activity in OGIS development has centered on a cartographic paradigm wherein geographic features may be defined by their spatial extent, attributes, and relationships. More recently, there has been increased focus on earth imaging applications and their unique requirements for spatial representation and description.

Environmental modeling is a discipline (or set of disciplines) that should both inform and be informed by the OGIS process. Certainly, the process of environmental modeling may be improved by direct access to salient geographic information for iterative analysis, compared to typical disjoint parameterization approaches where a geodataset is used to set boundary conditions for a model.

Traditional geoprocessing functions, such as spatial selection, thematic overlay, and measurement, may be incorporated directly into a model, along with emerging capabilities like spatial analysis and statistics. At the same time, modeling will require new interfaces to geodata that may not be incorporated into the specification presently.

OGIS will also have to expand its scope with respect to both three-dimensional data and temporal data. Although the basic geometric primitives for volumes and solids are in place, these are incomplete. The criticality of temporal data is recognized, and the draft specification notes that the spatial types and interfaces must become spatiotemporal types and interfaces. Inclusion of an environmental modeling application testbed in the OGIS process would go a considerable way in refining these issues and making true interoperability between geographic information systems and modeling systems.

Conclusions

Open GIS, distributed object technologies, and the Open GeoData Interoperability Specification provide a model for access to wide-area heterogeneous data and applications. The OGIS architecture allows application developers and users to identify, evaluate and utilize geographic resources - including spatial datasets, geoprocessing tools, and models and procedures - without being constrained by the distinctions among data organizations and processing environments. It will enable application domains not traditionally included in the field of GIS, such as environmental or process modeling, to interoperate with geodata stores and geoprocessing services, and will facilitate access to a broader range of modeling functionality to the GIS user. The OGIS project will result in detailed specifications for system components, plus a testbed for prototyping software and data models.

References

British Columbia Surveys and Resource Mapping Branch (1993), Spatial Archive and Interchange Format: Formal Definition, Release 3.0, December 1993. Province of British Columbia, Victoria (BC).

Object Management Group, Object Management Architecture Guide (1992). Object Management Group and X/Open, Cambridge (MA).

Object Management Group, Common Object Request Broker: Architecture and Specification, (1991). Object Management Group and X/Open, Cambridge (MA).

OGIS Project Technical Committee, The OpenGIS Guide, Introduction to Interoperable Geoprocessing (1996), Kurt Buehler and Lance Mckee, ed. Open GIS Consortium, Wayland (MA).

Open GIS Consortium (1996), The Open Geodata Interoperability Specification, Draft Base Document, December 22, 1994; rev January 8, 1996. Open GIS Consortium, Wayland (MA).

US National Institute of Standards and Technology, Spatial Data Transfer Standard (SDTS) (1992). Computer Systems Laboratory, National Institute of Standards and Technology, Gaithersburg (MD). Federal information processing standards publication 173.

GIS Research Specialist, REGIS
University of California
CEDR - 390 Wurster Hall
Berkeley, CA 94720

Tel: +01 (510) 642-9205

Fax: +01 (510) 643-5571

gardels@regis.berkeley.edu

B.Beattie, F.Coenen, A.Hough, T.J.M.Bench-Capon, B.M.Diaz and M.J.R.Shave

Spatial Reasoning for Environmental Impact Assessment

Abstract:

The potential of Geographic Information Systems has yet to be fully realized. Currently they are severely limited in that they do not provide for any spatial reasoning capability concerning the data they contain. This is partly because GIS data representations are not directly compatible with the predicate logic representations used by existing approaches to spatio-temporal reasoning.

There are many applications where a reasoning capability would offer signific ant benefits, such as the provision of support for the drawing up of environmental impact assessment plans with respect to the siting of factories, noise and air pollution control and effects on flora and fauna. The use of quad tesseral addressing to represent the GIS entities allows space in two, three and higher dimensions to be ``linearized'' so that established one-dimensional constraint-based reasoning techniques can be applied. This avoids the combinato rial explosion of directional relations which occurs between entities in two and higher dimensions, as we are able to express relations in terms of ``before'', ``equals'' and ``after''.

Furthermore, conventional raster and vector GIS map ea sily onto this data representation. The technique has been built into a spatial reasoning system, the SPARTA (SPAtial Reasoning using Tesseral Addressing) system which is described in the paper.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Ranjan Muttiah, Raghavan Srinivasan, Bernard Engel

Development and Application of Neural Network Interface for GRASS GIS

Abstract

The theory and development of a public domain neural network package for the GRASS GIS system is described. Classical classifiers using Bayes' selection rule, nearest means, and nearest neighbor were included in the interface for comparison against neural network predictions. Sample application of the package is presented erosion best management practices (BMP) identification, and remote sensing.

Introduction

Neural networks are a computational method of data analysis that are an extension of traditional statistical methods such as regressions (White, 1989), and function approximation (Baum and Haussler, 1988; Hertz et al., 1991). In statistical regressions the modeler has to *a priori* specify the functional form of the relationship likely to exist in the data set (nonlinear vs linear vs multiple regressions). The best functional form for the data is based on an error measure such as the least squares criterion. Neural networks, form an "internal weight" representation of the data as to minimize an error criterion (usually least squares) without too much *a priori* judgements about on the functional form for the data (McClelland et al., 1986). This for example provides a means of automating classification of very large datasets. Since GIS systems are data intensive in the spatial domain, and many different types of datasets (remote sensing, topography, hydrography) are used to make decisions and judgements, neural networks may find a useful role in capturing expertise, and in interpolating and extrapolating knowledge as an aid to decision making. The GIS system chosen here was the Geographical Resources Analysis Support System (GRASS) developed by the US Army Corp of Engineers (CERL, 1993). The primary reason was that GRASS is public domain, and the model developer can write specific routines using the C programming language (Kernighan and Ritchie, 1984) and the GRASS GIS graphic libraries.

Neural Networks

Neural networks in their broadest sense could be defined as a collection of interconnected simple computational units that work together cooperatively to solve linear and nonlinear problems. The network consists of a section that receives input information (input units) from the problem domain, an internal weight structure (hidden units), and an output section (output units). The input units are connected to the output units by way of hidden units. In circumstances of linear relationships, the input units can be directly connected to the output units. Information to the network can consist of either input and output pairs (input,output) as in the case of back-propagation supervised learning, or just input as in unsupervised learning (Kohonen networks). The hidden units capture the non-linearity in the mapping between the input and output information. The neural network is first trained on sample data, and the internal weights are adjusted to learn patterns and trends in the data. Once trained, the network is used to predict on input data. If there are many more hidden units (free parameters) than there are data available, the network may not be able to generalize (extrapolate), and learning of the network may be hindered by the noise and measurement error in the data.

The neural network interface that was developed here for the GRASS GIS platform incorporated only supervised learning. We selected the back-propagation algorithm of McClelland et al., (1986), and Baffes (1989) and the quick-propagation algorithm of Fahlman (1988). In the back-propagation algorithm, the network is iterated in "weight space" to minimize the mean square error measure at the output nodes given by:

$$\eta = \sum_{j=1}^{n} [d_j - o_j]^2$$

Where dj and oj are the desired and actual values at the output units of the network. The actual values at the output units are calculated by propagation of the input information through the network (using scalar products of weights and inputs in each interconnection). The weights between the units is stored are in a weight matrix W. Each hidden unit sums its input, and then applies a transfer function to this sum. The commonly used sigmoidal transfer function is given by:

$$f(x) = \frac{1}{1 + e^{-ax+b}}$$

Where a is the gain or scaling factor and b is the bias or the amount of translation of the sigmoidal transfer function on the x-axis. It has been shown that back-propagation networks are equivalent to Fourier series approximation when these sigmoidal units are used (Lapedes and Farber, 1987).

In a fixed topology network like back-propagation, the number of hidden units have to be decided before training on the data. The weight update between hidden unit i and the output node j on the n+1 th iteration is given by the gradient descent:

$$\Delta w_{ij}(n+1) = \nu \delta_{p_{ij}} O_{p_i} + \alpha w_{ij}$$

Where $\delta_{p_{ij}}$ is the back-propagated error to the hidden unit j from the output units on presentation of input pattern p to the network, ν is a learning rate, Opi is the input reaching hidden unit j from unit i not in the same layer, and α is a momentum constant that uses the previous weight values to avoid local minima on the error surface.

In quick-propagation the weight updates are made according to:

$$\Delta w(n+1) = \frac{S(n+1)}{S(n) - S(n+1)} \Delta w(n)$$

Where $S(n+1) = \frac{\partial E}{\partial w(n+1)}$ is the error change on weight change between any two units in the network.

Classifiers in the NN Interface

To compare and evaluate the predictions made by neural networks, additional classifiers from the pattern recognition field were incorporated into the neural network interface. Since GRASS already has the maximum liklihood classifier (which assumes a normal distribution function for the training data), classifiers using nearest means, nearest neighbors, and Bayes' rules were developed. The nearest means classifier calculates the mean vector of each training class, and classifies by the minimum Euclidean distance between input and mean vectors. In the nearest neighbor classification, the covariance matrix of the training data Σ is calculated, and inverted using LU decomposition. If the covariance matrix is singular, the error is reported to the user and the interface returns to the main menu. On successful inversion of the covariance matrix, the input vector is then multiplied according to the distance formula:

$$d_i(X, Y^i) = \left[(X, Y^i) \Sigma_i^{-1} (X, Y^i) \right]^{1/2}$$

The mean distances from the input vector to all the training class vectors is then calculated (dk-NN). The first nearest neighbor (d1-NN) is taken as the input class to which the input vector belongs.

In Bayes' classifier, the misclassification error (overlap error of the class probability density functions) for the classes is minimized using the Fisher's criterion. The general classification rule is given by:

$$h(Z) = V^T Z + V_0 \underset{<}{>} , Z \in \begin{matrix} \omega_1 \\ \omega_2 \end{matrix}$$

Where

$$V = \left[\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_2\right]^{-1} (M_1 - M_2)$$

and V_0 is given by:

$$V_0 = \frac{(\Sigma_1)^2 V^T M_{1z} + (1 - s)(\Sigma_2)^2 V^T M_{2z}}{s(\Sigma_1)^2 + (1 - s)(\Sigma_2)^2}$$

In Fisher's criterion, s is set at 0.5.

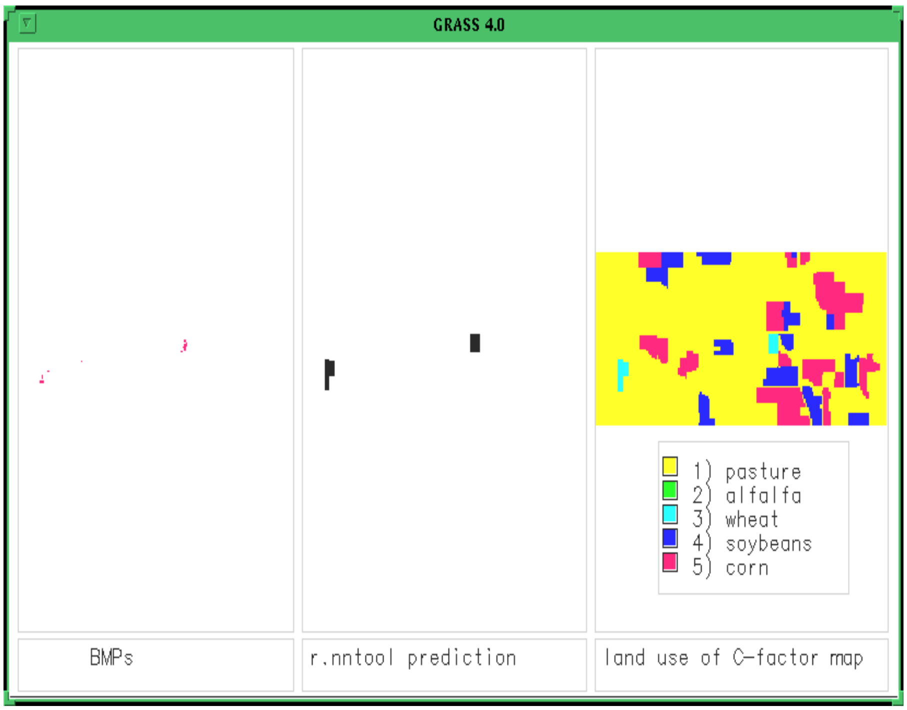
Implementation Details

The GIS interface was structured along the lines of the i.maxlik maximum liklihood classifier presently available in GRASS. A few new features were added. The user has the option of either using a pre-existing map with digitized training areas, or of selecting training areas using the interface (points, circles, polygons). Training areas that need to be deleted can be removed within the interface. The interface stores the training data for each class as a separate map layer. In the use of the tool, the user is asked to enter the name of the output map layer, the number of output classes, and the names of the input map layers. Using the lump option of the menu, the tool selects the "dominant" category within a specified window and generates a new map layer. The user can reset the resolution to the newly specified window size, or retain the old resolution in which he entered the tool. In existing GRASS routines, when resolution (window) of a region is enlarged, the middle pixel of the window in the lower resolution is selected. Training areas are selected using the define areas option. Using the zoom option, the user can zoom out to parts of the output map in which he wishes to delineate training areas. If classification is using two input vectors, the user can view the scatter plot of the training data, and selectively remove outlier points or points that cause conflicts (i.e., same input vectors belonging to two different classes). The training and input are also written as ASCII files for further data exploration outside of GRASS (such as in the use of the public domain xgobi viewer of Buja et al., 1986) If using remotely sensed data, the neural network interface allows input of spectral bands and training data as in the i.points program of GRASS. Histograms of the training data can be generated. After calculating the covariance matrix of the training data, the interface prints out the eigenvalues of the matrix. The eigenvalues could be used to identify the dominant or import features of the input map layers used in classification (see Fukunaga, 1972).

Land Management Application

The land management application presented here consists of those areas requiring best management practices since they have soil losses above the soil erosion tolerance limit (areas requiring best management BMP practices), and those land areas that have soil losses below soil loss tolerance. The Indian Pine watershed north of the Wabash river in West Lafayette, Indiana was selected as the study area. The USLE K, LS, C, and P maps for the study area were used to calculate the soil loss from each cell in GRASS using the USLE equation

(R*K*LS*C*P) (Wischmeier and Smith, 1978). The soil erodibility K-factor map was obtained from the K-factor in the Natural Resources Conservation Service Soils-5 database. The slope-length LS-factor map was obtained by running the r.watershed program in GRASS which determines the LS-factor based on the elevation map. The cropping-management C factor map was obtained by assigning values based upon the observed crop rotation practices within the Indian Pine watershed (mainly corn-soybeans). The agricultural fields were digitized from aerial photographs maintained by the Agricultural Conservation and Stabilization Service (ASCS) in Lafayette, Indiana. The conservation practice P-factor map was obtained by assigning management practice values upon observation of the fields in the study area. The rainfall and runoff erosivity index R was obtained from the USDA erosion losses hand book (Wischmeier and Smith, 1978). Those areas of the study area above tolerable soil loss (USLE T) requiring best management practices are shown as dark areas in the first column of Figure 1 (please click on the image for better view).



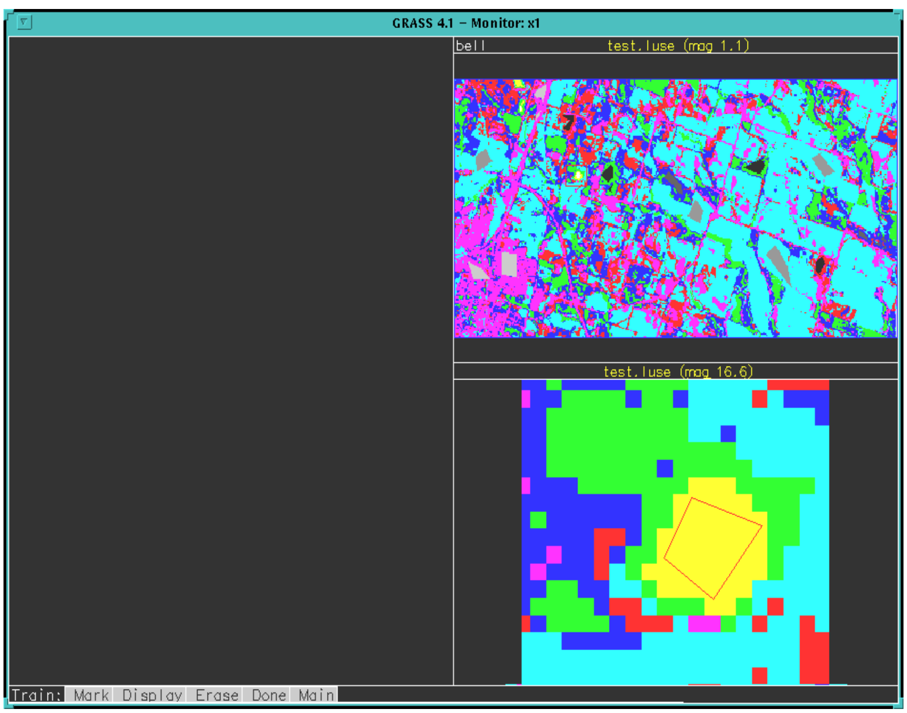
The best management practices (BMP) areas are clustered into two areas. Training data for r.nntool was selected from the lower left corner of the BMP map. The input data to the neural network (we chose to use quickprop) consisted of the USLE factors, and the output units consisted of binary data by pixels representing whether an area required BMPs or not.

The output of the neural network tool (r.nntool) after training is shown in the second column of Figure 1. The dark areas represent the areas requiring BMPs. As shown, the neural network has predicted whole field areas as requiring BMPS. The areas displayed correspond exactly to the shape and size of fields as shown in the C-factor map in the third column of Figure 1. All the field areas predicted as requiring BMPs also had points within them that had soil loss above the tolerance limit. This shows that the way the data was represented and presented by pixels to the neural network will cause prediction of more global features (Note: this would also be true of classical classifiers).

Remote Sensing Application

Neural networks have found many interesting uses in remote sensing because they allow integration of remote sensing and other complementary landuse information in image classification. Classical classifiers, such as maximum likelihood and nearest neighbor classifiers, have been primarily applicable with only satellite image band information. Neural networks allow for linear and non-linear mappings between satellite spectral data, complementary landuse information (eg., land ownership,slope and aspect), and landuse classes.

A thematic mapper (TM) composite for Temple, Texas using the second, third, and seventh channels was used to identify land use categories using ERDAS and ARC/INFO (McKinney, 1993). The TM scene was taken on March 14, 1992 (used here by permission of Nature Conservancy, Austin, Texas). The TM data was rectified using the road map of Temple, Texas. Areas were identified as either water, forests, rangeland, agricultural land (cropland and pasture), and other (urban, barren, and other categories). The TM composite was then imported into GRASS. To confirm the classification, we mounted a Trimble GPS unit called Pathfinder Basic+ (Trimble, 1991) on a vehicle and drove around the area of the TM coverage of Temple. The GPS readings were done non-differentially. Non-differentially, the GPS unit can be precise to 30 meters. Field GPS surveys were made at two different rangeland sites, five different agricultural land sites, one water body site, six different urban sites, and one forest site. There was agreement of observed landuse to those predicted from the TM composite, except for the water body in the northwest corner of the image which had a smaller coverage than that of the TM image. For application of the neural network tool, a smaller area (57 square kilometers) of the Temple area was subset from the larger scene.



(figure 2)

Input into the neural network consisted of the visible green band (.52-.60 micro meters) and the mid-infrared band (2.08-2.35 micro meters). The quick propagation network was chosen and training sites were selected for water, agricultural land, range land, forests, and other categories. Figure 2 shows the selected training sites (the sites for each training class are stored as separate map layers). The scatter plot of the selected training sites are shown in

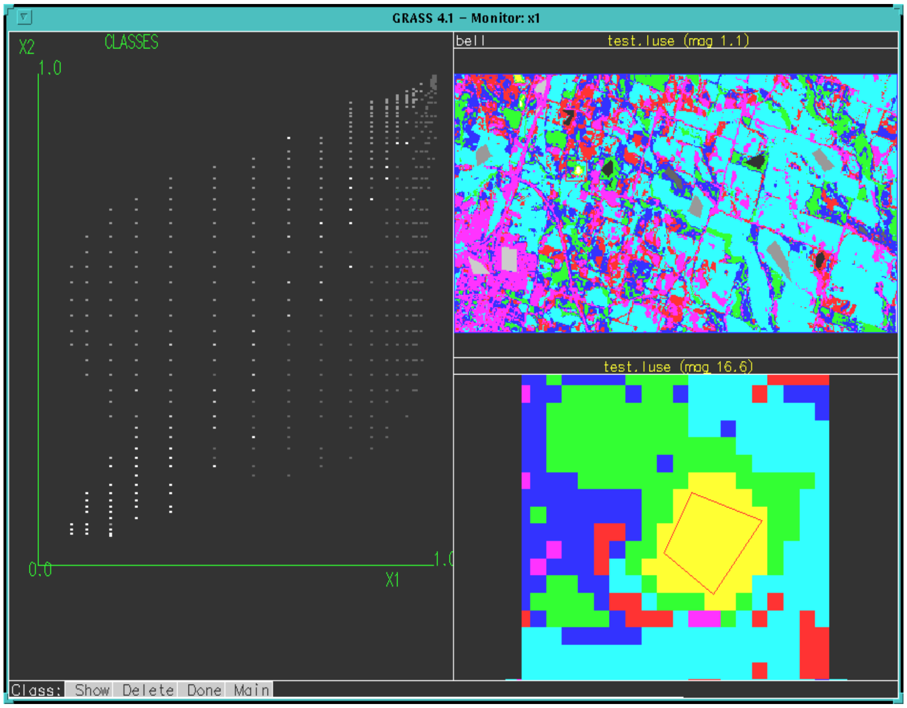
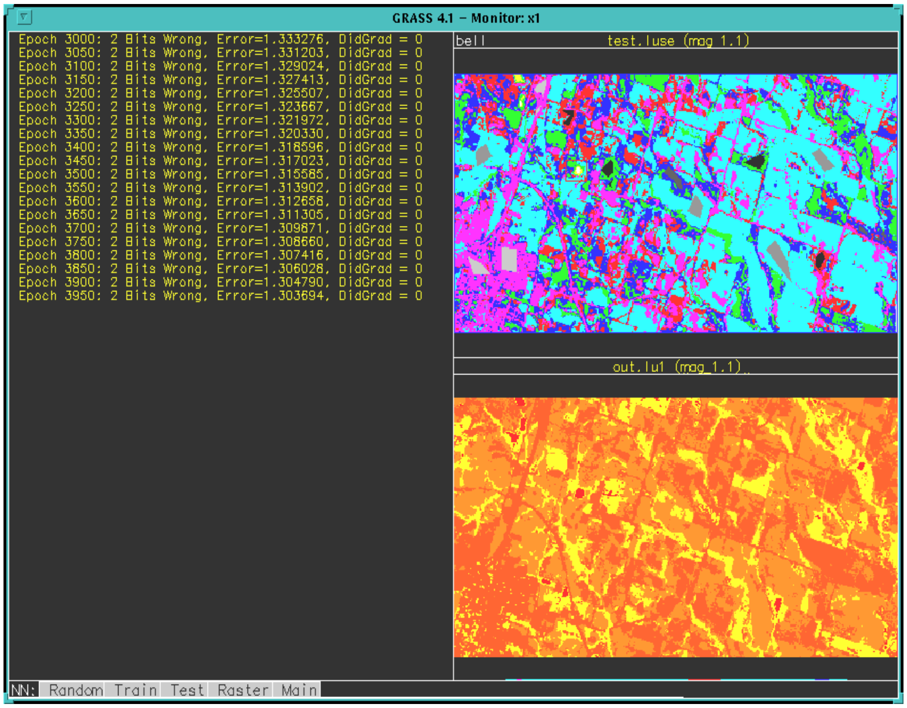


figure 3

Data were interactively cleaned where inappropriate training sites were selected, and where there were overlap of the wrong pixel classes in the selected training areas. Figure 3 shows the error at the output units of the quick propogation network on the training cycles.



The network converged to a mean square error of 1.30 after 4000 iterations (epochs). Twelve hidden units were used (based on some trial and error), and the network had two input units, and 5 output units. There were a total of 177 training points (water had 7 training points, forest had 92 points, agricultural land had 51 points, range land had 8 points, and other had 19 points). Once the network had converged, testing was done and an output raster file then generated and displayed (lower right corner of the above Figure 4). A nearest means classification was also performed on the data from an option available in r.nntool Comparisons by area are shown in Table 1. Table 1. Comparisons by landuses for the study area.

Land Class	Composite (%)	neural networks (%)	nearest means (%)
Water	0.11	0.65	1.88
Forest	9.36	6.84	9.89
Agland	44.41	43.37	28.32
Range	16.61	10.23	22.32
Other	29.51	38.91	37.58

The neural network predicted agricultural areas rather well, but under predicted rangeland significantly (probably because of the smaller sample size), over predicted other landuses, and somewhat under predicted forested areas. Depending on the accuracy to which a user wishes to obtain results, it is debatable whether neural networks would give much more than traditional classifiers when using only spectal band information (neural networks will learn all the data presented to it, so purity of the data training elements is important for good learning). This fact probably accounts for the poor prediction of the "other" category which has the barren and urban areas grouped together.

Summary and Conclusions

A neural network tool has been developed for the GRASS GIS platform, and has been shown to make predictions close to land surveys and linear classifiers in the remote sensing application. The land management application showed how training on clustered data can yield to more coherent global features. The r.nntool program has classical classifiers built into it for comparisons with neural network predicted maps. The program is public domain and users wishing to obtain a copy can contact the author.

References

Baffes, P., 1989, NETS Back-Propagation ver. 2.0: Software Technology Branch, NASA, Johnson Space Center, Houston, TX.

Baum E.B and D. Haussler. 1989. "What Size Net Gives Valid Generalization ?" In: NIPS I. Ed. D.S. Touretzky. Morgan Publishers, 2929 Campus Drive, San Mateo, CA 94403. pp. 81-90.

Buja, A., C. Hurley, and J.A. MacDonald. 1986. "A Data Viewer for Multivariate Data." Computer Science and Statistics: Proc. 18th symposium on the Interface. Am. Stat. Assoc. Washington D.C.

U.S. Army Corp of Engineers(CERL), 1993, GRASS Users Manual, ver. 4.0: Construction Engineering Research Laboratory. Champaign, IL.

Fahlman, S., 1988, Faster Learning Variations on Back Propagation: An Empirical Study: in Proceedings of the 1988 Connectionset Models Summer School.

Fukunaga K. Introduction to statistical pattern recognition. Academic Press Inc., Boston, MA. 1972.

Hertz J.A., A.S. Krogh, and A. Palmer 1991, Introduction to the Theory of Neural Computation: Addison-Wesley Pub. Co., Redwood City, CA.

Kernighan, B.W., and D.M. Ritchie. 1984. The C programming Language: Prentice-Hall Inc, Englewood Cliffs, NJ.

Lapedes A., and R. Farber. 1987. Nonlinear Signal Processing Neural Networks: Prediction and System Modeling. Technical Report: LA-UR-87-2662. Los Alamos National Laboratory, Los Alamos, NM 87545.

McKinney, T. 1993, Landuse Map of Temple, Texas: School of Forestry, Texas A&M Univ., College Station, TX.

McClelland J.L., D.E. Rumelhart, and the PDP Research Group. 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. The MIT Press. Cambridge, MA.

Trimble Corporation, 1991, GPS-Pathfinder Basic Users Manual: Trimble Navigation, Surveying and Mapping Div., Sunnyvale, CA.

Wischmeier W.H., and D.D. Smith. 1978. Predicting Rainfall Erosion Losses -- A guide to Conservation Planning. USDA. Agric. Handbook No. 537. 58 p.

White H. "Learning in Artificial Neural Networks: A statistical Perspective." 1989. Neural Computation. \fB1\fP, 425-464. MIT Press, Cambridge, MA.

Author Information

Name: Ranjan S. Muttiah
Title: Agricultural Engineer and Associate Research Scientist
Organization: Texas Agricultural Experiment Station, Blackland Res. Ctr., Temple, Texas 76502
Telephone: 817-770-6659
Fax: 817-770-6678
Email: muttiah@brcsun0.tamu.edu

Name: Raghavan Srinivasan
Title: Agricultural Engineer and Associate Research Scientist
Organization: Texas Agricultural Experiment Station,Blackland Res. Ctr.,Temple, Texas 76502
Telephone: 817-770-6670
Fax: 817-770-6678
Email: srin@brcsun0.tamu.edu

Name: Bernard Engel
Title: Associate Professor Organization: Agricultural and Biological Engineering Department, Purdue University, West Lafayette, Indiana 47907
Telephone: 317-494-1198
Email: engelb@pasture.ecn.purdue.edu

USING ARTIFICIAL NEURAL NETWORKS FOR PREDICTION OF SOIL CARBON DYNAMICS

Daniel Kimes and Elissa Levine
NASA/Goddard Space Flight Center, Code 923, Biospheric Sciences Branch,
Greenbelt, Maryland 20771 USA

The feasibility of using artificial neural networks (ANNs) as a method to model soil organiccarbon dynamics was tested. ANNs represent a versatile tool which is well suited for analysis of soil characterization data. They are designed to learn patterns or relationships in data from being given a set of inputs, and to recognize complex non-linear interactions within soils data that may be otherwise difficult to determine. ANNs also have the ability to generalize or abstract results from imperfect data and are insensitive to minor variations in input (e.g., noise in the data, missing data, or a few incorrect values) which may be present in soil data.

ANNs of various types were trained and tested on their ability to predict soil organic carbon using the USDA NRCS individual pedon data base and the aggregated STATSGO data base. Results of the best trained networks at each scale were used to develop models specific to soil orders and suborders. These models can then be used to predict values of soil carbon not available from existing data, to improve our understanding of the role soils play in carbon dynamics, and to provide a prototype for using ANNs to model other soil properties.

Controlling a GIS by the expert system EXCEPT

Thomas Wachter-Harms
TUHH
Department of Urban Ecology
e-mail: wachter@tu-harburg.d400.de

Birgit Wendholt
IBM Germany, Development Section
e-mail: wendholt@tu-harburg.d400.de

EXCEPT, EXpert System for Computer-aided Environmental Planning Tasks, was developed by the Technical University of Hamburg, Department of Urban ecology, and IBM Germany from 1989-92. Based on the methodology of Environmental Impact Analysis, the expert system provides the technical means (1) to implement a large variety of assessment methods in knowledge bases, (2) to apply these to specific assessment issues and (3) to explain and document assessment results, processes and methods. The current project is concerned with the integration of EXCEPT and a GIS, since we have identified a number of correspondences between the data models and the procedures in both systems. We will enhance the EXCEPT data model with a hierarchy of spatial objects to establish the reference to geographic objects on the GIS side. Based on this extension we will show, how to modify single EXCEPT assessment steps to trigger the corresponding spatial GIS operation. Since we have integrated the application of spatial operations into the overall assessment process of EXCEPT, we are able to interpret results of GIS operations by help of an EXCEPT documentation that explains the method underlying the assessment process.

Jonathan Raper and David Livingstone

SPATIO-TEMPORAL INTERPOLATION IN FOUR DIMENSIONAL COASTAL PROCESS MODELS

ABSTRACT

The key 1990s challenge in environmental modelling has been to create models with sophisticated spatio-temporal representational structures matching the problem domain which can generate testable predictions about the functioning of environmental systems. Many researchers have considered geographic information systems (GIS) when coupled with environmental models to be suitable for this purpose and have adopted them in a wide range of studies (Goodchild, Parks and Steyaert 1993). However, it is clear that the use of GIS places constraints on the representational scope of the coupled system: GIS are two dimensional, layer based, geometry-indexed systems and are often difficult to link to existing environmental models as Livingstone and Raper (1994) have shown. In summary, GIS are often difficult to couple to environmental models, and, if this is achieved the representational compromises required are prejudicial to the overall aims of environmental modelling.

This scenario seems to require some new thinking on the design of environmental modelling systems. Raper and Livingstone (1995) have argued that the design of an integrated system should be driven by the nature of the environmental system and its spatio-temporal structure and set out the design of an object-oriented geomorphological modelling system called OOgeomorph. Underlying the OOgeomorph design is a computer-aided software engineering (CASE) environment capable of building data models which directly implement entities and functional or spatio-temporal relationships derived from any source theory. OOgeomorph was used to represent the May and Tanner (1973) theory of coastal evolution in an integrated 4D modelling framework.

This paper has two main aims: firstly, to briefly describe the OOgeomorph model and its implementation for a coastal geomorphological theory; and, secondly to explore the implications of the OOgeomorph design and its representation of space and time for populating models with data and carrying out subsequent analysis.

USING OOGEOMORPH IN ENVIRONMENTAL MODELLING

The design of OOgeomorph

The driving philosophy behind the development of OOgeomorph has been to create a system that is suitable for formulating and testing theories within a geomorphological research context. Two fundamental points were realised (Raper and Livingstone 1994), firstly that the system needed to be capable of handling spatio-temporal representations and secondly that the system should not impose a restrictive, proprietary data model. These observations resulted in a number of general design approaches to be adopted in the implementation of OOgeomorph. These can be summarised as follows:

Design Approach 1

The separation of proposed geomorphological models from the observed or derived data that is used to populate these models.

This approach is consistent with the 'layered' philosophy of the Universal Geographic Information eXecutive (UGIX) system design proposed by Raper and Bundock (1993) which ensures that the low-level data structures of the stored data do not exert any influence on the structure of the geomorphological representation. A conventional GIS or other spatial database, the Geomorphological Spatial Database or GSD (figure 1), can therefore be used as a repository for observed data required by OOgeomorph. The GSD could be any form of spatial database ranging in sophistication from ASCII files of x,y,z coordinates to a fully topologically structured GIS.

The data in the GIS is 'mapped' onto the models under test by the creation of two class structures. Firstly, a class structure called 'geomorph_system' which represents a testable model; and secondly, a class structure called 'geomorph_info' which implements a generic representation of the observed data. The classes in 'geomorph_info' can be regarded as metadata which describes the associated data and encapsulate the translation mechanisms required to extract it from the GSD. The 'geomorph_info' classes are therefore a layer that separates a high-level geomorphological data model from a low level GIS data model and are decomposed according to the form of the data rather than the form of the geomorphological representation.

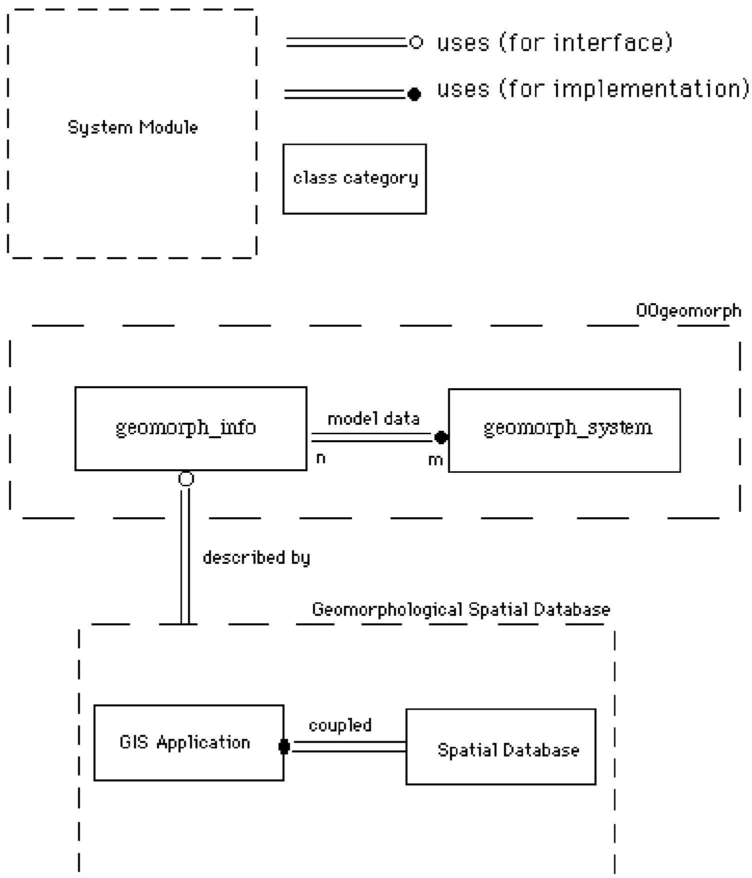


Figure 1 System Architecture - source Raper and Livingstone (1995)

Design Approach 2

<u>The structure of the data model should be organised according to geomorphological concepts.</u>

Since there is no way of knowing in advance what organising concepts a particular geomorphologist will select then the structure of the data model needs to be flexible. This resulted in the adoption of what could be termed a customisable CASE approach to enable the creation of a model structure which follows a set of user-defined principles. These user-defined principles are equated by the authors to the adoption of a 'meta-theory' which determines the nature of a set of 'basic level categories' to be used as a starting point for the construction of the geomorphological data model. In the rest of the paper these 'basic level categories' will be be considered to be processes, forms and materials in accordance to some kind of geomorphological 'meta-theory'. In a full implementation of OOgeomorph it would be possible to define a different configuration of 'basic level categories' before creating the data model by defining instances of the class 'geomorph_system' - this, however, is not discussed any further here.

The 'basic level categories' can be used by a geomorphologist to create specific sets of 'process', 'form' and 'material' classes relevant to the investigation or environment, the idea being that a number of different class structures can be investigated without having to redefine the underlying spatial database. Theories or conventions can be used to determine which variables or 'observables' are relevant to the geomorphological phenomena under investigation and the range of values allowed for each used as a criteria for 'geomorphological object-creation'. In OOgeomorph a geomorphological object is known as a 'phenomenon instance', in that it is an instance of an aggregated class made up of collected observables of the process, form and material classes. This approach ensures that geomorphological phenomena are not forced into a taxonomic hierarchy (as in some GIS designs) based upon geometric descriptions of their state at a particular point in space-time, but rather that they are assembled from the spatio-temporal coincidence of relevant process, form and material observables according to user-defined criteria.

Design Approach 3

<u>Geomorphological observables should be referenced to four dimensions and be capable of overlapping with each other.</u>

Each observable, such as wave approach angle or surface elevation, is valid over a spatio-temporal region determined by the conditions under which the measurements were collected or variables defined. Initially a convention has been adopted whereby this region is decomposed into its spatial and temporal components; a 'time of knowing' and a 'location of knowing'. Both these references and the value of the observable being measured constitute its attributes. These attributes are derived from the data described in 'geomorph_info' and may be in a variety of different spatial data formats. The idea behind the creation of the 'time of knowing' and 'place of knowing' attribute is to make the spatio-temporal information reconcilable. The next step is to define a phenomenon class as an aggregation of relevant attributes, organised according to specified observables, each of which 'know' where they are derived from and have associated methods to enable geographical processing tasks to be performed according to the functionality of the GSD. Any additional processing tasks required would need to be added to OOgeomorph.

The concept of time currently adopted in OOgeomorph equates to the concept of 'valid time' considered by Worboys (1994) and others. The 'time of knowing' is a point in time (recorded as a clock time and date) when each observable of process, form and material can be known. This idea needs to be extended so as to acknowledge that measurements have a estimable 'range of applicability' prior to modelling or analysis. This extension might be in the form of extending the point into a one-dimensional bounded region or by the addition of an associated validity method eg a probability distribution. Any phenomenon instance is therefore a collection of process, form or material observables (figure 2) each with its own space-time reference which may be disjunct, overlapping or identical.

In a typical scenario phenomenon class may be defined by the aggregation of <u>three</u> process observables, <u>two</u> form observables and <u>three</u> material observables (i.e. eight in all). If these observables themselves have a value, a 'time of knowing' and a 'place of knowing', then there will be at least 24 process, form or material attributes associated with a PI. The spatio-temporal structure of each PI and the set of PIs in the phenomenon class is entirely in the hands of the geomorphologist, since each of the eight process, form and material observables can have a different spatio-temporal extent. This design means that PI's are spatially and temporally heterogeneous ranging from highly observation-dependent forms ('over this space at this time') to infinite steady forms ('always, everywhere'). An implication of this design is that PI's of widely differing 'scales' are created and compared rendering the concept much less important. The key control over scale is, therefore, the granularity of the theory governing selection of process, form and material observables.

It is an explicit design aim for OOgeomorph that such complex expressions be facilitated as it permits the storage together of geomorphological phenomena considered comparable as a working hypothesis, though they may be both spatially and temporally disjunct and have differing values for their process, form or material observables. Any differences can then be examined using tools created to compare the PI's under objective conditions.

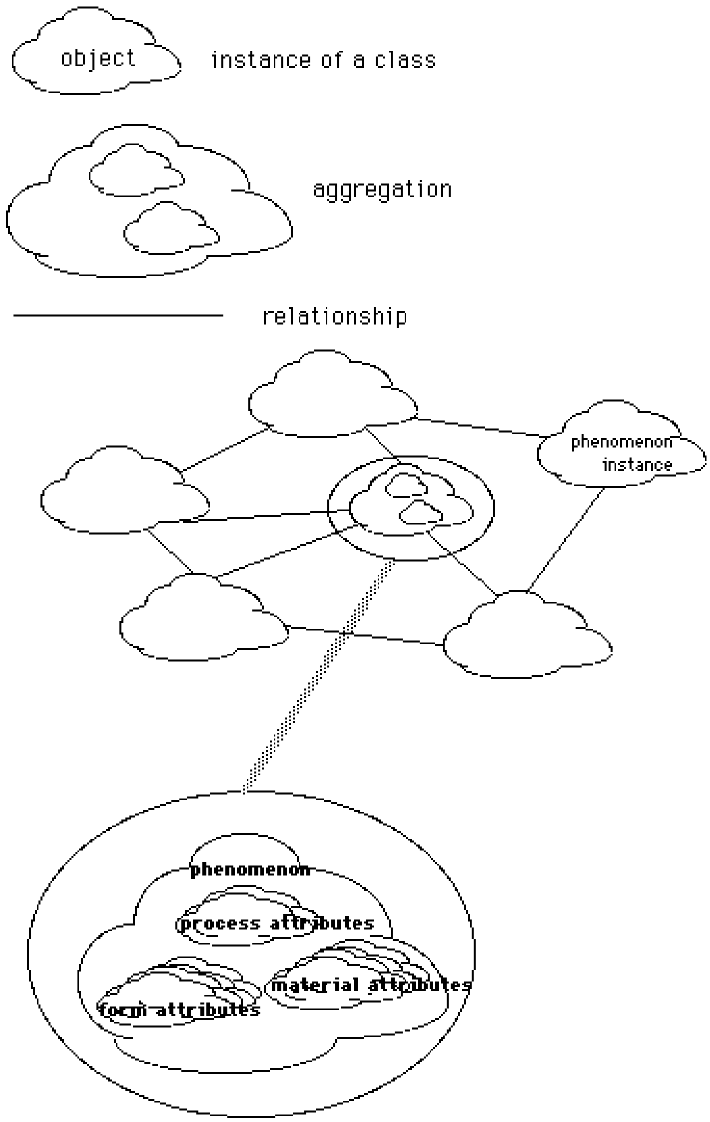


Figure 2 Structure of Phenomenon Instances (PI's) - source Raper and Livingstone (1995)

Design Approach 4

To permit the assignment of behaviour, expressed as mathematical models, to data about geomorphological phenomena.

Given that a phenomenon instance (PI) is likely to be spatially and temporally heterogeneous it is necessary to implement a set of tools to operate on PI's. Such tools can be implemented as operations 'encapsulated' with the phenomenon sub-class. An important class of these tools involves interpolation. Many proprietary GIS have spatial interpolation procedures and, as long as the GSD permits external command processing all of these are available to OOgeomorph either before or after the creation of the PI's. Since PI's are objects existing in a spatio-temporal framework then if one is to create integral 4D objects to represent dynamic geomorphological phenomena, eg coastal spits, that can be related to other PI's then a set of formal 4D interpolation operations are required. The definition of interpolation operations is just part of a wider set of tasks including the specification of 4D object-object relations and the identification of characteristic 4D forms so that a 4D object language can be used to formalise 4D operations.

Operations relevant to OOgeomorph include:

classification operations to form subsets of the PI's based on the identification of sub-structures;

generation of new PI's by creating new observables according to mathematical/statistical models;

temporal and/or spatial interpolation to harmonise the space-time bounds of observables;

temporal and/or spatial generalisation when PI's vary slowly over space and time.

Using OOgeomorph in a coastal study

To indicate the use of the OOgeomorph design an example is given here that is based upon the May and Tanner theory of coastal cell development. It has been created to test some of the hypotheses about the spatio-temporal development of coastal cells which Carter (1988) advanced, in this case with specific reference to medium term coastal development.

The first stage is to identify the concepts in the theory and which of the 'basic-level categories' they belong to (Table 1).

Form Category	**Process Category**	**Material Category**
Internal points	Wave energy E	Unconsolidated sediment
Erosion	Wave crest approach angle a	
Deposition	Longshore wave power PL	
High spring tide level	Sediment discharge Qs	
Low spring tide level		

Table 1 Concepts used in the coastal cell theory as formulated by Carter (1988) - source Raper and Livingstone (1995)

The next stage requires that the observables, consisting of values or types with discrete four dimensional spatio-temporal referencing, used to define or model these concepts be defined as classes in one of the 'basic level categories' (Table 2). The key issue in this procedure is the granularity of the representation i.e. the spatio-temporal region over which these observables are valid. In most cases the theory should provide a definition of the granularity of representation such that the temporal and spatial domains can be identified.

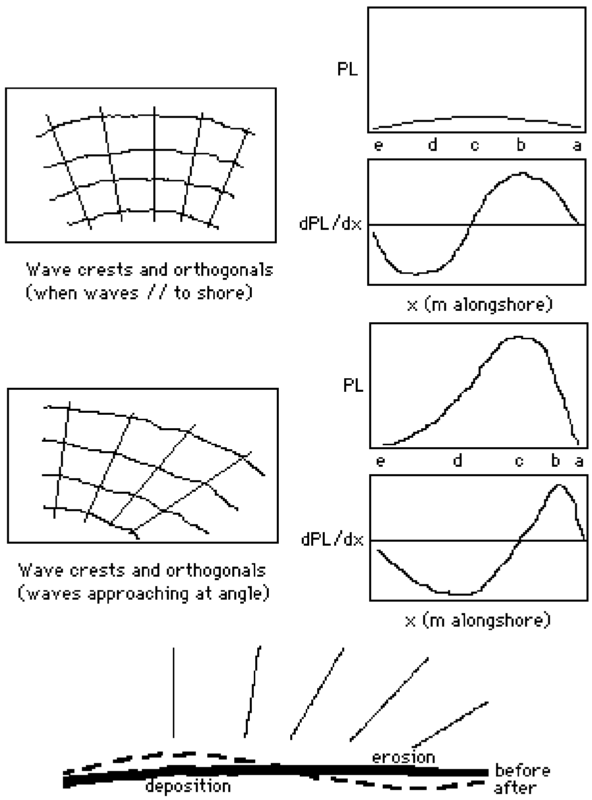


Figure 3 Basic relationships of the May and Tanner theory (a-e are 'internal points')

In this case the May and Tanner theory provides concepts of granularity in the form of the 'internal points' of the coastal cell, which themselves represent minima and maxima of longshore wave energy power flux within the cell. To implement this representation it was decided use measurable shoreline cross-sections at-a-time the spacing and frequency of measurement defining the granularity of the study. Shoreline cross-sections are normal to wave run-up and tidal fluxes and therefore offer versatile 'candidates' for the internal points. They also offer appropriate ways to discretise the other concepts in the theory such as sediment flux, angle of approach for wave crests and material properties.

Form observables	**Process observables**	**Material observables**
Candidate point	Wave crest approach angle a	Sediment type
High spring tide level	Longshore wave energy power PL	
Low spring tide level	Sediment discharge Qs	

Table 2 Form, process and material observables for the coastal cell representation in OOgeomorph - source Raper and Livingstone (1995)

Observables in table 2 have an internal structure in the form of the following attributes:

Value of attribute	{Char * }	an alphanumeric value for the attribute
Time-of-knowing	{Char * }	clock time and date
Place-of-knowing	{Geom * }	a series of points defining a shoreline cross-section

The phenomenon class - a data model of a coastal cell , will have 21 'attributes' organised into sets of three and a phenomenon instance will consist of many observable instances. In order to actually test the validity of a coastal cell model and to parameterise the model from empirical observation requires the extension of the model to include methods associated with either the observable classes or the phenomenon class - suitable spatio-temporal interpolation methods are an important requirement for this part of the model. For example:

Determining erosion and deposition changes between measuring cycles

Since the cross sections may not always be taken in the same place nearest neighbour cross-sections in space over time need to be identified. A simple solution is to select nearest subsequent time then nearest spatial distance. However it is entirely possible that the cross-section from a slightly later survey but in a more similar spatial location may be more suitable - if suitable units can be derived eg from consideration of controlling processes, then a 4D distance can be used with dimension [L][T]. [L] in distance units [T] in time units.

The result of this operation is to create a new, derived, observable which can be added to the definition of the phenomenon class if required i.e. net shoreline change.

Identifying internal points 'a' to 'e'

Since it is difficult to determine longshore wave energy power (PL) directly (Carter 1988), this operation requires secondary evidence to identify candidate points. Approaches include those based upon wave refraction modelling and beach elevation/sediment type variation alongshore. Candidate points can also be proposed from evidence of erosion and deposition recorded in the previously created net shoreline change observable.

Defining the boundaries of features such as spits

One of the main aims behind the design of OOgeomorph is to enable the definition of features such as spits including statements about their boundaries. Features such as spits are modelled as phenomenon instances (PI's) or sub-sets of the observable instances that make up a PI. The boundary of a spit over space and either at or over a particular period of time is derived from the cartesian product of all the time and places of knowing of the observables that make up the spit. Such a boundary is a set of possibly spatially and temporally disjoint regions, it is proposed that an initial spatio-temporal interpolation would define the 4D equivalent of a convex hull around these regions. The state of the spit boundary at a particular time would then be approximated by a 3D cross-section.

The result of this operation would be a new phenomenon instance, possibly constituting a sub-set of another PI. It allows the geomorphologist to define the boundaries of features such as spits by using morphometric analysis of slope angles and other observable values and to use information such as tide levels as delimiters. The geometry resulting from such an operation is associated with a phenomenon instance and allows vectors of movement to be computed between successive states of a spit's boundary.

The structuring created by OOgeomorph makes it possible to execute four dimensional 'range queries' which look for space-time coincidences. In studying the evolution of coastal cells, typical questions which take the form of a range query include:

> Over what longshore distance does an internal point move on average during a predetermined period?
>
> Where do rates of movement for feature boundaries defined by an operation on the PI's reach their maximum and when?
>
> When do the candidates for internal points differ most in longshore terms when defined by differing criteria?

In order to efficiently and consistently perform such range queries as well as to define spatio-temporal relationships between objects the authors are working on the formalisation of 4D operations, including the process of spatio-temporal interpolation.

All the above operations and queries are designed to be carried out on any set of phenomenon instances defined in the study zone. The existence of this kind of system will make it possible to carry out such queries on many different phenomenon instance sets defined using different tools. The authors have been collecting the data required to instantiate this representation for the Scolt Head barrier island (in North Norfolk, England) at regular intervals over the last five years and are engaged in the implementation of this data model for this field site.

THE SPATIO-TEMPORAL STRUCTURE OF ENVIRONMENTAL MODELS

OOgeomorph has been designed to allow modellers considerable flexibility in the way that observables are spatially and temporally referenced. This was considered necessary since the variables in many environmental models are spatially and temporally heterogeneous and each may discretised from a continuous external reality differently. Kemp (1993) considered this problem at length and argued that the solution was to express all variables in (discrete) field form such that all modelled variables could be determined for the same locations at the same successive times. Kemp suggested that the process of expressing the variables in field form should, therefore, be documented and encapsulated with the datasets at the time of creation. However, she notes that 'this information must be deduced and appended by the modeller him- or herself who, it is hoped understands at least a little about the nature and sampling of the phenomenon being represented' (p121). OOgeomorph has been designed, firstly, to permit a wider range of discretisations than those involving the expression of all variables over a single field associated with a single time, and, secondly to make the assignment of variables to spatial and temporal ranges fully explicit within the modelling process.

In OOgeomorph each observable has its own spatio-temporal attributes viz. location (x, x,y or x,y,z sets) and time-of-knowing (year/month/day & hours/seconds). The spatial referencing may be one, two or three dimensional in nature and describes the position at which the observable's value is measured. The temporal referencing is a one dimensional point reference to the time at which the observable's value is known. Hence, as a minimum, the OOgeomorph design requires that the observable's value plus its location and time be known or stated as an assumption, and stored as attributes of the observable class. Instances of observables may not necessarily overlap in time and space if they are measured in the field rather than being co-located by assumption or by implication of other models. To illustrate the different design approach taken by OOgeomorph the spatial and temporal structure of a series of typical environmental modelling problems are considered below for two of the model types distinguished by Burrough (1996).

Firstly, 'rule-based models' relate states of variables using logic and set operations e.g. IF salt marsh elevation is greater than or equal to spring tide level THEN the marine sedimentation rate equals zero. In such an example the resulting state is assumed to hold for all locations where the initial state is true. The range of meaning can be extended by using fuzzy set membership values. In a conventional environmental model linked to a GIS the known marsh elevations would be discretised as a field; any values in the field greater than the height of the spring tide would be selected and written out as a new field corresponding to null marine sedimentation. In OOgeomorph 'elevation' would be implemented as an 'observable' with attributes of location (x,y,z) and time-of-knowing while spring tide level would have attributes of location (z) and time-of-knowing (in this case predicted days and times). The logical model would be implemented as a method in OOgeomorph and would generate a new attribute of null marine sedimentation for any elevation point higher than the value of tide level as a query. Note that by making the calculation of the level of spring tide a method, the OOgeomorph approach could also generate output for any particular spring tide level without creating a new field. OOgeomorph would also not require that scattered values of elevation be converted into a discrete field-based surface model to implement the model.

Secondly, deterministic physical (or mathematical) models relate variables through mathematical relationships e.g. the May and Tanner (1973) theory of coastal cell evolution used in Raper and Livingstone (1995) which relates wave crest convergence at the shoreline to the distribution of wave power along the shoreline. By taking the first derivatives of the rate of change of longshore wave power, the points of maximum and minimum wave power can be determined, and through them the rates and locations of erosion and deposition along the shoreline can be calculated. In a conventional environmental model linked to a GIS, the wave convergence and the distribution of wave power would have to be calculated outside the GIS for an idealised x axis corresponding to the shoreline. By discretising the rate of change of longshore wave power at a user-specified interval, offsets from the shoreline of a particular time could be calculated. In a vector GIS the 'offsets' would be drawn as lines and connected to form a predicted future shoreline. In OOgeomorph observables for convergence, longshore wave power and shoreline would be given value, location (x) and time-of- knowing attributes. Methods to determine longshore wave power from convergence and erosion/deposition from rate of change of longshore wave power would operate on instances with a shared time and generate new instances of the shoreline observable at user defined x intervals. By taking this approach OOgeomorph allows the integration of all the calculations in the same system and can automatically carry out the calculations for any specified time. It would also be possible in OOgeomorph to creat an observable called 'actual shoreline' which could automatically be compared with a 'predicted shoreline' over any time interval.

In each of the above cases the comparison of the 'conventional' approach with the 'OOgeomorph' approach specifies an idealised procedure: in normal modelling practice the situation is often much less clear cut and many 'pre-processing' operations are required to harmonise the spatial and temporal limits, resolution and symmetry of the discretisation of the variables. For example, some variables will be sampled at a point and the value generalised to be representative of an area e.g. records from a gauging station (climatic or geomorphic processes). Conversely, some variables will be sampled simultaneously over an area creating a field of values which are summarised to generate a single point (e.g. remote sensing of wave height). Similarly, sparse temporal measurements may need to be generalised to cover unmeasured periods (e.g. elevation surveys), or, samples may be taken at frequent intervals leading to the need to summarise values (e.g. recording of water levels using a chart recorder). The spatial and temporal heterogeneity may also involve an assymetry in summarising or generalising operations: hence, there may be spatial constraints on the operations in a particular direction or temporal constraints over a particular period

SPATIO-TEMPORAL INTERPOLATION USING OOGEOMORPH

Facilities for spatio-temporal 'interpolation' (an expression used here to cover summarising and generalising operations) should, therefore, be built into environmental modelling systems and tools designed to document their use. It is suggested here that one way to manage the spatial and temporal heterogeneity of the variables is to record the 'range of applicability' of both the location and time-of- knowing attributes by creating using two additional attributes. Spatially, this 'range of applicability' could be a radius around a point, a buffer around a line or polygon, a vertical range or a resolution change limit for a field. Temporally, the 'range of applicability' could be a period of time which is symmetrical or assymmetrical around the temporal point.

Such 'range of applicability' attributes when added to the basic spatial and temporal attributes would greatly facilitate a variety of forms of spatio-temporal 'interpolation' in OOgeomorph. Firstly, by using simple queries the 'range of applicability' for a set of variables could be explored to check their comparability in spatial and temporal domains. If the set has a single outlier in range terms, methods can be constructed for that variable to recalculate summarisation or generalisation operations on the raw data so as to harmonise the range with the other variables aggregated into the phenomenon class. Similarly, if the set of 'ranges of applicability' for a group of observables being entered into an algebraic expression is highly heterogeneous in nature then methods can be developed to define a commensurable spatial and temporal range for all the observables.

Spatio-temporal interpolation can also be applied to the domain of the variable instances as well as the spatial and temporal domains of the variables themselves. When variables can be considered to be evidence (either singly or jointly) for the existence of some geoscientific phenomenon the spatio- temporal distribution of the instances may be of critical importance. At present few if any modelling systems can represent, manipulate or visualise such data. The design of OOgeomorph makes it possible to pose four dimensional range queries to determine the bounding limits of any phenomenon. Such limits correspond to a four dimensional envelope or 'hypercuboid'. In the case of a study of the rate and style of change of coastal forms the configuration of such an envelope is of considerable interest since its four dimensional form and structure may be correlated with the energy inputs to the system.

However, the structure of the available instances may be deficient in some respect. For example, certain observations may be missing or there may be errors associated with them. In this case it may be necessary to interpolate observations lying spatially and temporally 'between' observations at known locations. OOgeomorph can be used to locate space-times where there are no observations and the nearest observations to the 'empty' area. Currently the authors are developing heuristics for the interpolation of points for coastal surveys that it was not possible to carry out for logistical reasons.

CONCLUSIONS

The key implication of the research carried out in this paper is that by adopting a new spatial database design that assigns four dimensional coordinates to all 'observables', a range of environmental modelling operations can be carried out within one system without the need for the low level coupling of a model and a GIS. Chief amongst these operations are the ability to establish the 'range of applicability' of variables that are to be related in a formal modelling statement and the ability to interpolate gaps in spatio-temporal observation data.

Implementation work on the OOgeomorph system is under way within the framework of a UK Ministry of Agriculture, Fisheries and Food (Coastal and Flood Defence Division) research project on coastal spits and nesses. The resulting system will be used to manage field collected and simulation- generated data on the spatio-temporal behaviour of these landforms and to enable the testing of hypotheses on the driving processes.

REFERENCES

Kemp, K.K (1993) Environmental modelling with GIS: a strategy for dealing with spatial continuity. National Center for Geographic Information and Analysis, Santa Barbara, USA. Technical Report 93-3.

Burrough, P.B. (1996) Opportunities and limitations of GIS-based modeling of solute transport at the regional scale. ASA-CSSA-SSSA Bouyoucos Conference, Mission Inn, Riverside, CA, USA, May 1-3 1995, 'Applications of GIS to the modelling of non-point source pollutants in the vadose zone'.

Raper, J.F. and Livingstone, D (1995) Development of a geomorphological data model using object- oriented design. International Journal of Geographical Information Systems 9 (4), 359-83.

Carter, R.W.G. (1988) Coastal environments : an introduction to the physical ecological and cultural systems of coastlines. Academic Press, London.

May, J.P. and Tanner, W.F. (1973) The littoral power gradient and shoreline changes. In Coates, D.R. (ed.) Coastal geomorphology, Binghampton State University, New York.

Livingstone, D.E. and Raper,J.F. (1994) Modelling environmental systems with GIS: theoretical barriers to progress. In Innovations in GIS, (ed. Worboys, M.), London, Taylor and Francis, 229-240.

Goodchild, M.F., Parks, B.O. and Steyaert, L.T. (1993) Environmental modelling with GIS. Oxford, OUP.

Raper, J.F. and Bundock, M.S. (1993) Development of a generic spatial language interface for GIS. In Mather, P.M. (ed.) Geographical Information Handling- Research and Applications, Wiley, Chichester, pp113-143.

Worboys, M.F. (1994) Unifying the spatial and temporal components of spatial information. Advances in GIS research (Proceedings, 6th International Symposium on Spatial Data Handling, Edinburgh, 5-9/9/94), 505-17.

Jonathan Raper# and David Livingstone*

#Department of Geography, Birkbeck College, Univ. of London 7-15 Gresse Street, London, W1P 2LL, UK

tel. +44 171 631 6470 fax. +44 171-631 6498 j.raper@geog.bbk.ac.uk

*School of Geography, Kingston University, Penrhyn Rd, Kingston, KT1 2EE, UK

tel. +44 181 547 2000 [2021] d.livingstone@kingston.ac.uk

An GIS-based Many-Region Disaster Preparedness Model for the United States

Sam Cole,

National Center for Geographic Information and Analysis,

University at Buffalo

Paper to be presented at the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, USA,

January 21-25, 1996. Sam Cole

An GIS-based Many-Region Disaster Preparedness Model for the United States

ABSTRACT

This paper describes the construction, solution, and application of a many-region county-level social accounting matrix of the United States to be used to assess economic damage arising from natural disasters, such as floods, hurricanes, or earthquakes, (including economic earthquakes), and to improve the viability and efficiency of subsequent recovery efforts. Since the full empirical model is extremely large, concepts drawn from Geographic Information Systems are used to aggregate and view transactions by geography and sector so as to focus the model on the target locality and its immediate linkages. Thus the system is organized as a "virtual model" whereby the many-region accounts are condensed to a set of basic data associated with each locality and a set of estimated parameters which allow selected local accounts to be re-constructed and solved as required. Digitized national data sets for states and counties, are used to estimate county-level social accounts and inter-county spatial interactions. The approach may be extended to sub-county localities, such as urban neighborhoods. As a complete set of transactions between the 3000-plus counties the United States, the model may be used to explore how major disasters propagate through the national economy.

Creating such a system in a GIS-like environment presents a major challenge. In principle, every locality described in the model is connected to every other through structure, time, and space - changing any variable in one region affects every variable in all other regions. Even though specific sectorial locality to locality impacts may be small (say, the impact of a small earthquake in Santa Barbara on Buffalo, New York), the aggregate effects across regions of the nation (say, the Northeast) can be substantial. Most GIS deal well only with variations in structure, and do not deal well with time varying and place-to-place flows. Moreover, since many data are missing or must be estimated there remains a question of how the virtual model, and the metadata that describes it, should be organized. Last, for real-world applications, the system should be manipulated via a relatively straightforward Decision Support System (DSS) interface. There are thus several critical choices to be made, and the

paper describes these choices and their rationale.

Section One explains the approach which is directed initially at estimating medium-term economic and social costs and losses to populations and institutions in small localities, where ready-made planning instruments and proactive contingency plans may not be available. Section Two describes the construction of the social accounts, to include estimates of economic transactions within and between counties and regions, as well as details of economic distribution across households, businesses, and lifelines. Section Three describes the solution procedures using a time-dependent approach for conceptualizing economic processes. This offers consistency between the construction and solution of the many-region social accounts, and allows the magnitude and time-scale of failures arising from natural disasters and subsequent reconstruction efforts to be integrated into the solution of the model in situations where damage from a disaster is spatially and temporally distributed across infrastructure and delivery systems of several localities. Finally, Section Four considers the appropriate DSS, drawing a distinction between controlled research applications and the chaotic circumstances of a natural disaster.

INTRODUCTION

The rationale for the specification of the modeling system being developed in this paper, arises from the following considerations.

In many cases, natural disasters have their most severe impacts on isolated localities and small and impoverished communities. In the United States natural disasters are a national problem, experienced at the local level (Berke and Beatley, 1992). For example, a major disaster like the "500-year" Midwest floods of 1993 had minimal impact at the national, or even at the state level, yet many small communities may never be able to recover. As a fraction of annual income or even normal business cycle swings, even Hurricane Andrew had modest impact on Dade County, Florida - yet South Dade County and Homestead were devastated. This alone demands that models of relatively small districts be made available.

Disasters often impact most severely on poor and marginal populations. In many cases, this can be traced to poverty and inappropriate development, such as inferior infrastructure or housing, or limited insurance and other defensive resources (Cuny, 1983). For similar reasons, disasters tend to impact small businesses more severely than large enterprises. Ideally, a model should describe the situation of these groups and activities explicitly, as well as their links with the wider economy and community.

Even when they not impacted directly, people and businesses may be affected through damage to lifelines such as water supply or roads, or through indirect effects such as the loss of livelihood or markets. Even in aggregate, the indirect effects on a community are often far larger than the direct effects (Eguchi et al, 1992). Therefore, models need to be economy wide, so as to account for all sectors of an economy and all segments of a population (see e.g. NRC, 1990).

In general, the smaller a community, the greater will be the importance of its linkages to neighboring districts, and the more vulnerable it will be to damage to lifelines, such as power and water supplies, or transport and other communications. Moreover, the smaller a community the greater will be the spill-over effects to other communities, the more likely it is that neighbors also will be impacted directly by the disaster, or that feedback effects through their economies will be important. Thus, models must have an explicit account of a locality's links to its neighbors and the world beyond, and in many cases models must be multi-regional.

Disasters affect populations for an extended period of time. Most disasters and recovery activities operate on a variety of time scales that are characteristic of the physical, social, economic, and technological systems involved. As far as possible, it is necessary to represent these within a model, through the manner in which the impacts are calculated. Ultimately, the analysis of disasters presents an almost intractable network problem. Even so, one vital aspect of recovery programs or pro-active measures is to recognize that the vulnerability of the locality to disasters can be reduced though the application of various systems principles to the physical, economic, and social aspects of disaster preparedness (Kameda and Shinozuka, 1989).

Models need to be situated within the overall development process. While the immediate priority must be to attend to the life-threatening consequences of the disaster (such as medical and shelter needs), it is also necessary to plan for an economic recovery which provides an opportunity to improve the quality of development if the hardship from future disasters is to be reduced. In the past, victims and places often have been disadvantaged even further by deficient recovery programs. For this reason, disasters are best viewed as part of the development process, providing opportunities, as well as tragedy (Jones, 1989). A model based on this perspective must bring together relevant social and economic categories, so that both the damage caused by the disaster, and the proposed recovery strategy, can be assessed in the context of the long-term goals and institutional structures of the community.

Since natural disasters typically take place with rather little specific warning, it must be possible to provide the relevant planning tools that are adaptable to the situation of any community and type of damage, and become available as soon as possible after the disaster, so that they can be used to evaluate alternative proposals for recovery, before irreversible commitments are made. Even after a disaster has occurred, there remains considerable uncertainty as to the extent of damage or the most appropriate recovery strategy, the model should be updatable and flexible, so as to respond to evolving community needs.

All of the above - the need to provide analysis quickly, to provide models for small localities, describing specific sectors or lifelines, and particular types of household or community - suggests a rather high level of empirical detail and technical sophistication. This conflicts with several practicalities - the availability of data, the understanding of complex systems, and the needs of the planning process. The last cannot be ignored since, when the separation between the modeling and policy making becomes too great, the modeling loses much of its potential use (United Nations, 1994). While many expert and other decision support system have been developed in an attempt to bridge this gap (see e.g. Batty and Yeh, 1991), they have yet to confront the empirical and institutional chaos and complexity of much disaster planning.

Even though the above is an incomplete list of the challenges for addressing the consequences of major natural disasters, it presents a formidable task for economic modeling. While the present project has faced most of these issues, the exercise described in this paper focuses largely on questions of model construction and application in the context of a specific disruption to a lifeline systems in a metropolitan region of the United States.

The underlying theoretical question being addressed here is how do perturbations to an economic system propagate through structure, time, and space, and how well can regional modeling capture this process. Geographic Information Systems offer one promising means for bringing together the various data indicated, and to link the decision support system and directly to other studies and data bases.

Modeling Economic Disasters

The social accounting method used in this paper is a variety of input-output model (see e.g. Pyatt and Roe, 1979), and widely recommended as a core empirical device for national, regional and local planning (United Nations, 1994). Input-output models have been applied at the national or regional level for disaster assessment, with some recent efforts at the county level (e.g. West and Lenze, 1993) and small territories and islands (Cole, 1993).

Social accounting models have the particular advantage that, given the requisite data, both the supply and the demand sides of the economy can be described as a network that can be mapped onto its physical and social counterparts. It was argued earlier that this is necessary to describe and evaluate the consequences of specific types of damage in a useful fashion. The potential contribution of the methods adopted in this paper is that it extends the possibilities for constructing detailed input-output type models for small localities, and for introducing fairly complex disaster and reconstruction scenarios, taking account of changes in the internal structure of the economy.

Most economic transactions depend directly on physical lifeline systems - for example, purchases of power and water by businesses and households, the trucking of goods between industrial areas and to markets, the flow of information within and without the region via telecommunication links. The impacts of earthquakes on lifeline systems involves not only earthquake resistant constructions of individual components but also system recovery with the aid of network redundancy, back-up facilities, and restoration work, that are to be followed by reconstruction and improvement for the future earthquake (Kameda and Shinozuka, 1989). Because the nodes in the input-output tables represent localized production and consumption activities and the links shows flows of goods and services, such tables are especially appropriate for representing production, exchange, and consumption activities.

In order to address the complexity of the consequences of natural disasters it is necessary to reformulate the assumptions used in standard input-output calculations. This is especially so when there is a complex of events arising from the partial failure of several activities, resulting in a more general failure of the economic network as a whole.

CONSTRUCTION OF THE MANY-REGION ACCOUNTS

The above considerations place a very demanding set of requirements for model construction, especially when it is recognized that data at a small spatial scale are often restricted (for confidentiality and other reasons), and information on the direct effects of disasters are usually incomplete and collected in an ad-hoc fashion. On the other hand, since the damage caused by a major natural disaster can affect the structure of an economy in dramatic ways (through the loss of entire industries or lifeline systems) so that even a model which captures the key features and linkages within an area's economy, before and after a disaster, and allows the broad outlines of a reconstruction strategy to be developed, can be useful. To this extent the requirements on precision for the model may be somewhat less than for a conventional economic impact calculations. Thus, the initial goal has been to make possible the rapid construction of a first-cut social accounts based impact model for any locality within the United States, using readily available data, while providing for the subsequent improvement and extension of the model, to sub-county localities. (Figure 2) (Figure 3)

Development of the Model

In order to develop the requisite analytic procedures, it has been necessary to construct a detailed county-level many-region model of the entire United States. While this may appear a somewhat convoluted procedure, it provides the parameters necessary to estimate models for smaller, sub-county localities, and transactions between counties across the entire United States, providing indirect estimates of information that are otherwise not readily available.

The basic techniques used to construct the present model are relatively straightforward. Total supply and demand for each commodity and factor of production are estimated for every county and by scaling the accounts in the United States table. These supply-demand imbalances are then used to estimate the parameters of a spatial allocation model for each activity, and to provide bilateral inter-regional trade matrices. These flows are combined with the scaled regional matrices to give the many-region social accounts. Finally, the accounts are aggregated so as to highlight the locality of immediate interest within an overall matrix of manageable size. The main principles of the approach described here have been piloted in earlier phases of the project (NCEER, 1993). Whereas these pilot studies required considerable "hands-on" treatment using some specially collected data, the present approach is largely automatic or mechanical using national data sets available in digitized form (e.g. on CD-ROM) with data handling and presentation manipulated through geographic information system (GIS) techniques, and using generalizable algorithms to scale and solve the models (Miller and Blair, 1985).

Representation as a Virtual Model

The principal difficulties with extending the techniques to the many-region United States model arise in management of large volumes of data: a model of the transactions between the 3000 counties of the United States, each with up to 20 activities per region could requires a matrix with 36x108 entries (a table covering several footballs fields!). The size of this matrix is increased several fold when the results of the model, the corresponding changes in transactions for the years following a disaster event are also inventoried. Even though many of the transactions are zero, this would still require a vast amount of data, To deal with this the system is organized as a "virtual model" whereby the many-region accounts are condensed to a set of basic data associated with each locality and a set of estimated parameters which allow selected local accounts to be re-constructed and solved as required. Typically, only a small part of the total United States model is required in full detail at any one time - i.e. that describing the locality impacted by a disaster and its neighbors. The rest of the model may be aggregated according to the distance and direction from the target locality, for example, along major radial routes from the area. Thus, the system is designed to focus in on the segment of the matrix that describes the disaster area, in much the same way that a GIS system allows us to zoom a particular geography. A key task has been to devise appropriate estimation, and aggregation rules for this process.

Overview of Empirical Implementation

The empirical implementation of model construction and application is carried out through a series of computer programs. The procedure falls into several steps, each comprising a group of programs; data preparation, model estimation, model assembly, and model solution. The assumptions described below allow these steps to be carried out recursively. More elaborate scaling might require this part of the calculation to be iterated.

During the initial data preparation, information is extracted from the national data base which is carried on two CD-ROMs. Geographic data (county coordinates) and economic data (i.e. selected county level data required for matrix scaling) are transferred to separate state files (AreaCoo and AreaData respectively). These data are then used to construct a data file focused on the states in which the target counties are clustered (AreaBloc), and a second file (BlocDist) giving the bilateral distances (or alternative measures of impedance) between. A similar file is prepared for the national level, comprising states and clusters of states, or sub-county data. County or sub-state clusters for are then combined to a single file (for example, Mississippi, Tennessee, and Arkansas, as well as state data are combined in the Memphis file). This file (BlocData) has the same output format as those prepared at the county level (AreaBloc) in order that following steps are unchanged. The national social accounts (US-SAM) are scaled to the regions using data from the BlocData files in order to estimate domestic demand and supply for all regions (BlocScale). Equivalent information comprise the data base for the virtual model.

These estimates and the bilateral distances then are used in the calibration program (BlocGrav) to derive the bilateral transaction matrices for selected regions. A balancing program (BlocRAS, not shown) can be used at this stage. For the assembly of the final accounts, data from BlocGrav (or BlocRAS) are aggregated to final regions and sectors (BlocAgg), and the corresponding domestic data are prepared (BlocComb). These data are then combined into the final table using BlocSAM. Last, the model is solved (BlocImp) either to provide characteristic multipliers for the target region, of by introducing data on specific events (BlocEvent). A master program to input data and control program flow from within the GIS window is being developed.

Regional Supply and Demand

The supply and demand for commodities and factors by region and sector are the core data set for the virtual model - that is they provide the information from which the local models are extracted as required. Supply and demand are estimated by scaling a total requirements table for each county or regional bloc in turn scaled from a national table for the United States. The data required to build local area tables are not generally available (because they are too expensive to collect, or because they cannot be disclosed legally). The underlying assumption behind this scaling is that technologies (measured in terms of total factor and commodity inputs) across similar production sectors are uniform nationwide. A similar assumption is used for household consumption propensities.

The overall procedure for constructing the model includes two steps which involve the spatial aggregation of the model. The first is carried out prior to the calibration of the model parameters simply to reduce the number of entities involved in the calibration. Typically this procedure includes about one hundred clusters made up of single counties, within-state groups of counties, individual states, and multi-state blocs, such that the level of aggregation increases with distance from the target region. Collectively, these blocs comprise the entire United States. With this approach, the blocs of counties and states may be superimposed on the radial transportation networks which are ubiquitous in US cities, or modified to correspond to other systems, so that the regional clusters correspond to nodes on the lifeline network. The selection of the target county, and neighbors, is automated via a GIS interface.

Inter-regional Transactions

As there are no data on commodity flows between small localities, these too must be imputed. The estimation of inter-regional flows rests on the tremendous variation in the size, sectorial composition, and spatial disposition of county level economies in the United States. All localities exhibit an imbalance between amount of goods and service they produce and the amount they consume: excess supply from each locality is exported to other districts, and abroad, and vice versa, workers commute from districts where there is an excess of labor supplied by households, and so on. These spatial patterns have complex underlying causes, and there are several theoretical explanations of why the intensity of transactions should tend to decline with increasing distance between actors. This includes the need to limit the adverse consequences of unexpected delays, the presumption that underlies the method used to solve the present model. Throughout the estimation, a principle concern is to retain sufficient empirical information to estimate the model (see e.g. Griffith, 1991) since it is well known that any aggregation scheme introduces difficulties and ambiguities in the estimation of even simple models, manifestations of ecological fallacy, modifiable areal unit type concerns (see e.g. Fotheringham and Wong, 1991).

The inter-regional transactions in the social accounting matrix may be estimated by fitting parameterized spatial interaction models to the previously-estimated supply and demand for each commodity and factors. Once a plausible estimation of the bilateral transactions between regions is obtained, and anomalous results dealt with, the matrix for each commodity is re-balanced (see e.g. Cole, 1994).

Focusing Locality Specific Models

The above procedures for scaling levels of supply and demand and inter-county flows provides accounts for some 40 county-level models and a comparable number of regional and county clusters. This is still a relatively large model, and most of the information is not needed for consideration of any particular county. A second aggregation is used to focus the model onto an individual county (e.g. that which is closest to the epicenter of an earthquake or has suffered the most severe damage), or a string of counties (e.g. along a particular lifeline or natural phenomenon, such as an earthquake fault or river). In effect, this procedure targets a particular segment of the county-by-county United States economic model, just as a GIS zooms on a selected area of a map. (Figure 1)

Refocussing the model to a new locality involves repeating this second aggregation, but does not require that the model be re-estimated. For the second aggregation, total demand commodity-by-production tables first are scaled to the new combinations of regions as described earlier, and the bilateral trade matrices are organized by commodity and region into a commodity-by-commodity trade matrix which is then combined with the domestic commodity-by-activity matrices. This provides a combined many-region commodity-by-activity matrix, with the domestic economy of the selected region shown in full detail.

MODEL SOLUTION AND APPLICATION

Distributed Disruptions, Transaction Costs, and Uncertainty

The technique for solving the model rests on a time-dependent approach for conceptualizing input-output tables. The first important notion here is that the network of activities in any economy sets up a 'round-by-round' process that distributes income throughout the economy (see e.g. Miller and Blair, 1985). Thus, a change anywhere in the economy is magnified and transmitted throughout the community (and in some measure, throughout the world). This is the basis of the multiplier effect, that underlies all input-output type calculations, and provides an especially useful way of conceptualizing the propagation of events through structure, time, and space. The second underlying idea is that all economic processes involve a characteristic transaction lag - reflecting the time taken to design, finance, transport or produce goods, or simply to adjust to new circumstances (see ten Raa, 1986; Cole, 1988). As a consequence of the complex system of delayed feedback, it contains the multiplier process in any economic network always takes some time to build up to its full effect.

The prevailing spatial structure of any economy is in part a reflection of this system of delays, and of the effort by economic actors to minimize costs arising from it. Practically, actors seek to avoid the consequences of unexpected delays. Even under normal circumstances some proportion of transactions will be delayed unacceptably, and suppliers and their customers reduce losses by maintaining buffer stocks, or concentrating their business in nearby markets. In the event of a disaster, the proportion of failed transactions is increased beyond the capability of the normal system to cushion the event. Formalizing this process allows the magnitude and time-scale of lifeline and sectorial failures, and mitigation, or reconstruction efforts, to be integrated into the solution of the social accounting matrix. This includes situations where damage from a disaster is spatially and temporally distributed across infrastructure and delivery systems. The method can be related to Shinozuka et al's (1994) formulation of the fragility curve for a linearly connected lifeline systems, and to the method of Bates (1994) for analysis of travel time reliability. (Figure 4)

Decision Support Systems versus Expert Systems

The events to be simulated include the following:

i) Simple Disasters (or recovery) where the changes in output, income, and employment directly and indirectly can be calculated using standard input-output techniques (i.e. requires information on the proportion of each transaction affected directly by the disaster).

ii) Complex Disasters where the composition or source of inputs to particular sectors or households are changed or constrained in a non-linear fashion (requires e.g. information on household trade-offs, and the criticality or capacity constraints on production).

iii) Recovery Scenarios involving reconstruction of the activity over some specified time period (requires e.g. information about additional delays in performing transactions because of lifeline failure, the time taken for the transaction to recover to its previous or some specified level, or new links to enhance the robustness of the economy against future disaster).

iv) Prioritized Recovery incorporating recovery into a broader development strategy involving many conflicting interests (may require additional information on discount rates and trade-offs to minimize losses, for example within a discounted cost-benefit analysis).

The way in which the social accounts and the event matrix are used to aid decisions depend on the precise application. It is useful to distinguish here between two types of application in "well defined" and "poorly defined" or "chaotic" situations. The first approximates to the circumstances of a region as a research laboratory for disaster preparedness. The second corresponds to the type of situation that the modeling system developed here ultimately is designed to contribute to. These distinctions dictate the type of expert system that might be usefully employed. emphasize the balance between quantitative and qualitative representation, and the degree to which problems are well or ill-defined, and this the extent to which they can be modeled in a mechanical fashion (see e.g. Batty and Yeh, 1991). The promise of expert systems based on concepts of artificial intelligence has proved hard to realize and increasingly, even for well defined situations, the literature increasingly speaks of decision support systems rather than expert systems. The approach of this project falls into the generally accepted definition of aiming to help a community do for itself what otherwise an experts might be asked to do, but this nevertheless leaves open many issues of system interface design.

REFERENCES

Bates J. 1994. The effect of travel time reliability on travel behavior, 41st North American Regional Science Association. Niagara Falls, Canada

Batty M., and T. Yeh. 1991. The promise of expert systems for urban modeling.

Berke P. and Beatley T. (1992) Planning for Earthquakes: Risk, Politics and Policy, Johns Hopkins University Press, Baltimore.

Cole S. 1988. The Delayed Impacts of Plant Closures in a Reformulated Leontief Model, Proceedings of the Regional Science Association, Vol 65, pp 135-149.

Cole S. 1992. A Lagrangian derivation of a general multi-proportional scaling algorithm, Regional Science and Urban Economics, Vol 22, pp 291-

Cole S. 1993. Cultural Accounting for Small Economies, Regional Studies, 27.2, pp 121-136.

Cuny F. 1983. Disasters and Development, Oxford University Press, Oxford.

Eguchi R., H. Sekigson, J. Wiggins. 1993. Estimation of Secondary Losses Associated with Lifeline Disruption, 40th. North American Meeting, Houston, Texas. Fotheringham S. and D. Wong. 1991. The modifiable areal unit problem in multivariate statistical analysis, Environment and Planning A. Vol 23, pp 1025-1044

Griffith, D. 1983. The Boundary problem in Spatial Statistical Analysis, Journal of Regional Science., 23, 377-387. Also see (1993). Advances in Theoretical and Applied Economics, London, Kluwer.

Jones, B. 1989. The Need for a Dynamic Approach to Planning for Reconstruction after Earthquakes, NCEER, University at Buffalo.

Miller, R. and P. Blair. 1985. Input-Output Analysis: Foundations and Extensions, Prentice Hall, New York. NCEER. 1993. A Social Accounting Matrix Approach to Disaster Preparedness and Recovery Planning, National Center for Earthquake Engineering Research, NCEER 92-0002, Buffalo. NRC. 1990. The Economic Consequences of a Catastrophic Earthquake, National Academy Press, Washington.

Pyatt G, and A. Roe. 1977. Social Accounting Matrices for Development Planning with Special Reference to Sri Lanka, Cambridge University Press,

London and World Bank, Washington.

Pyatt G, and A. Roe. 1977. Social Accounting Matrices for Development Planning with Special Reference to Sri Lanka, Cambridge University Press, London and World Bank, Washington.

ten Raa, T. 1986. Dynamic Input-Output Analysis with Distributed Activities, Review of Economic Statistics, Vol 68, pp 300-310.

United Nations. 1994. Planning for Sustainable Development: Guidelines for Developing Countries, Department for Development Support and Management, DDDSMS/DEPSD, New York.

West C. and D. Lenze. 1993. Modeling Natural Disaster and Recovery: An Impact Assessment of Regional Data and Impact Methodology in the Context of Hurricane Andrew, Bureau of Business and Economic Research, Gainsville, University of Florida.

ACKNOWLEDGMENTS

This work was supported in part by the National Center for Earthquake Engineering Research (NCEER) under National Science Foundation (NSF) Master Contract BCS90-25010 and New York State Science and Technology Foundation (NYSSTF) Grant Number NEC-91029. Travel support and software also have been provided the NSF grant to the National Center for Geographic Information and Analysis (NCGIA) SBR- 88-10917.

Sam Cole,

National Center for Geographic Information and Analysis,

Department of Geography,

Wilkeson Quadrangle,

Ellicott Complex,

University at Buffalo,

NY 14620.

Phone/Fax 716-837-3924.

Animation Query Language for the Visualization of Temporal Data

Bruce Rex and John Risch

Pacific Northwest National Laboratory
Richland, Washington

(db_rex@pnl.gov, js_risch@pnl.gov)

Abstract

The emphasis on temporal query has been to return state values of temporally-related attributes for some discrete moment in time. The requirements for temporal queries with respect to visualization have not been well elaborated in the literature (Okazaki, 1993). A standard grammar and syntax for temporal queries has yet to emerge, although several temporal query languages (TQLs) have been offered in the literature (Snodgrass, 1992). No query language has yet come forth that places special emphasis on queries that are capable of returning animations. This type of query is especially useful in the analysis of systems that have strong temporal dependence.

This paper will examine issues related to dealing with time in visual information systems (GIS and other domain-specific systems) with special emphasis on system design in support of temporal queries which return animations. Issues related to the handling of time in databases will be examined, as will concepts of time as a dimension (or dimensions). A prototype Animation Query Language (AQL) will be introduced outlining a grammar and syntax that should be intuitive and generic. Specific problems associated with high-dimensional data, screen representations and human-computer interface will also be addressed.

Introduction

Multimedia technology, especially the use of sound and animation, is becoming an increasingly useful and accepted tool in a wide range of application areas. Certain multimedia techniques can also be effective for facilitating spatial information analysis. In particular, animation can be used to clearly and intuitively convey temporal interrelationships among multitemporal spatial data. The specification and generation of multitemporal data animations, particularly in real-time, can be complex and computationally expensive, however. The effective performance of such operations necessitates the advanced computational power afforded by multiple processor hardware platforms.

Fortunately, desktop systems incorporating multiple processors and dedicated graphic coprocessors can be purchased today for only a few thousand dollars. Such computers can make use of large amounts of RAM to operate in parallel on megabytes of data and rapidly generate very high-quality graphics. These advanced hardware architectures enable, for the first time, a tight, near real-time coupling between data, data modeling systems, and visualization tools. Such high-performance hardware platforms, when coupled with advanced, object-oriented software architectures, can open up new ways of processing, visualizing, and interacting with spatio-temporal data. Further, the advent of other innovative hardware technologies, such as six degree-of-freedom pointing devices and stereoscopic displays, make feasible a progression beyond the traditional 2.5D spatial model currently used by most GISs today.

This paper describes a general framework for managing and visualizing multitemporal spatial information. Although higher dimensionalities are being addressed in the literature, this paper will limit it's discussion to techniques for managing data within the framework of the three standard Euclidean dimensions of space and a single dimension of time (although the multiple dimensionality of time and resulting implications on temporal query will also be addressed). This is the basic 4-dimensional (4D) spatio-temporal abstract model. This paper address issues concerning the design of information systems to support the visualization of spatio-temporal queries made on such a model.

Architectural Issues

The current trend in computer system architectures is toward high-speed, wide-band single processor and multiple processor hardware running multithreaded executable modules. These modules can be regarded as individual programmatic objects containing both code and data. There is growing acceptance and use of the dynamic linking of such objects as seen in the Windows world in Dynamic Link Libraries (DLL) and Object Linking and Embedding (OLE). The Common Object Request Broker Architecture (CORBA) is gaining acceptance in the UNIX world. This is especially useful in distributed computing where multiple systems are involved in the solution of a single problem. In reality, however, CORBA is simply a large scale model of how multiple processors within a single computer share and distribute code and data.

Multiple processor architectures allow for the simultaneous computation on the same or different data by single or multiple algorithms, depending on the topology employed. Parallel microprocessor architectures provide several benefits. First, operations can be executed more quickly on data because less data is being analyzed at once by any given processor. Alternately, multiple operations can be performed simultaneously on the same data. Because of these advantages, larger-scale problems can be analyzed than were previously tractable, and moderate-scale problems can be processed at near real-time speeds.

Parallel hardware architectures can enhance visual information systems functionality and performance in several ways. Such systems can perform a variety of operations simultaneously, such as database operations, model calculations, user interactions processing, and display generation. Depending on the system hardware and software architectures, it is now feasible to develop systems comprised of coupled databases, simulation models, and 3D visualization systems, capable of supporting highly interactive, real-time modeling. The successful implementation of such a system is, of course, highly dependent on the data model used for structuring spatial and temporal information.

DBMS and Temporal Data

Underlying any information system is some sort of data management scheme, usually some form of database management system. The most widely used is the relational database management system (RDBMS) which relies on tables of attributes organized by entity relations. These databases have been in use for many years and are well understood. The most common method for storing temporal data in a relational database is to maintain attribute values in tables sharing a common discrete time value. The storage requirements for a relational database are dramatically affected by temporal granularity because temporally-related attributes need to be explicitly stored for every point in time. Data concerning some region which has been stored for a multi-year span might become very unwieldy if the temporal resolution needs to be finer than daily. The problem is exacerbated by the number of temporally-dependent attributes.

Another problem with coarse-grained temporal data storage is indeterminacy (Snodgrass, 1992). This occurs when the value stored in the database is representative of an attribute over a period of time, but does not capture the exact granularity of some temporal event. An example of this might be hourly temperature values. The temperature stored is the temperature measured hourly, but fails to capture the minute-level fluctuations over the hour. The data is then indeterminate because the database granularity is expected to be hourly, but the exact measurement has minute-level resolution and many possible observations have been ignored.

Object oriented database management systems (OODBMS or "object bases") enjoy a distinct advantage over relational systems by their ability to allow a model or submodel to be directly grouped with the appropriate data via class design. The object-oriented concept of encapsulation permits attributes and methods (data and its requisite modeling process) to reside together in a class. This encapsulation provides good problem abstraction and enhances query performance, because database searches are typically less arduous than they would be in a relational system. However, object bases can suffer from the same maladies as relational databases with respect to indeterminacy and other granularity problems whenever data is stored in snapshot style. Object bases are typically less bulky than the relational databases because they are not inherently tabular.

The great advantage of using an object model rather than a relational model for highly temporal data is due to the ability to encapsulate submodels into the class structure. The value here is that temporal values of related attributes can be calculated at runtime by the class' methods (submodels) as opposed to needing to be stored a priori, and then retrieved at runtime. The current generation of relational database management systems are beginning to allow user defined external procedures (models) to be called from fourth- generation language (4GL) macros. This is certainly a step in the right direction, but is still somewhat limited due to dependence on relational tables for data retrieval in support of submodels. This is an important step down the path of progress in visual information system design because complex, domain-specific models which were formerly dealt with exogenously to the information system can now be integrated into the system directly. Such integration promises improved productivity because the analyst need not worry about porting output data from the submodels into a separate visualization package or visual information system during each analytic iteration.

System Design

Some areas where spatio-temporal models are typically needed and applied are in ecological modeling, urban analysis, economic modeling, transportation studies and other time-space analysis such as found in architecture and behavioral studies. Domain experts have tuned many models over the years to provide accurate modeling of systems or processes. These models have been written to run on specific platforms in specific languages. In order to provide a visual information system that supports detailed high quality analysis, expert models must be integrable to the system.

There are three fundamental methods for tying models together into an analytic framework. The first method involves the use of several separate programs, each executed independently and sequentially. This is typical of expert modeling domains. Several submodels are executed, usually with the output of one module being fed as input to the next. There is usually a feedback component, where the output from an advanced submodel run is fed as input back into a submodel executed earlier in the process. This is usually a completely manual operation which, while eventually yielding the desired final product, is very labor intensive. Additionally, a combination of commercial software and user-written code is involved. For example, data may be collected in EXCEL, exported to a flat file according the specific input requirements of an domain-expert's model code, then the output of that routine is then imported back into EXCEL and either charted in that program or exported to whatever format a visualization package may require.

The second method involves the use of some form of remote procedure call to a submodel. This method is employed in ARC/INFO's Advanced Macro Language (AML). System calls can be made to shell scripts or precompiled programs which can be executed outside of the main information system environment. This method can be very slow and cumbersome, especially if the main information system and the submodel do not share the same data store or if the submodels are slow running and the main information system is stopped while the submodel executes. When the main information system and the submodels do not share the same data store, then it is necessary for the user to execute data transfer operations between analytic steps. Such steps typically include the loading of several output files from the submodel into tables within the main information systems relational database. This data transfer and loading can usually be done with scripts and macros, and thus can be automated in most cases.

The third and preferred method is to incorporate the domain-specific model directly into the main information system using a common data store. In addition, it is preferred to have the submodel code reside as a class method in an object-oriented database. This approach requires more work up front by system designers and programmers, but allows the analyst to concentrate more on analysis and less on data management. This is very important to the analysis of highly temporal data. The object approach permits a higher level of data abstraction by analyst. The common data store provides seamless data input/output between the database and submodels. This approach also scales well to multiple processor architectures where submodels can be assigned to individual processors and access the shared data of a common data store. Again, larger models can be analyzed more quickly than previously possible on single processor systems.

Temporal Dimensionality and Display Time

Temporal queries become even more difficult when time has behavior or dimension other than its typical one-dimensional linear form. The snapshot approach taken by relational database queries tends to model time as discrete, discontinuous moments. There are situations when such a simplistic model may not be an accurate representation of time. Time exhibits properties that are sometimes nonlinear. For example, the snapshot approach implies that time is discontinuous and discrete, in that it is only sampled at intervals. Time may also be log-linear, multithreaded (Chrisman, 1993), multidimensional and even fractal, depending on the application and how exactly time is perceived.

Snodgrass (Snodgrass, 1992) has proposed that time is at least two dimensional, with database time (transaction time) and event time (valid time) being measured along orthogonal axes in 2D, and provides a good illustration of this. So, if time can be multidimensional, can it also be fractionally dimensional as well? One would suspect so, but it is beyond the scope of this paper to examine those theories beyond the implications to temporal query. A properly constructed temporal method residing in a class structure should be able to handle time in whatever user-defined dimension is necessary. Likewise, there should also exist translation methods within the class structure to properly map nonlinear time onto a single temporal dimension that can be used for display purposes.

Display time is determined by the granularity of the native system time. This time is determined primarily by refresh time of the display device, but can be effected by the system clock speed as well. For example if spatial attributes are being calculated at run time, the algorithm throughput of a submodel may exceed the refresh rate of the display device. So the granularity of display time can determined by the efficiency of submodels as well as hardware constraints. If the submodels are excessively slow, visualizations can be cached for later real-time review (more on that later). The notion of display time begs the question: if valid time and transaction time are orthogonal in 2D, is display time orthogonal to valid and transaction time in 3D?

For static queries (queries based on a single, discrete time interval), display time is only relevant in that it takes some small amount of time to reconstruct the display image. For queries returning animations, however, there are several factors to consider. The speed at which frames are delivered to the display device must be taken into account. NTSC television refreshes at about 30 frames-per-second (fps). While computer monitor refresh rates vary considerably they are all faster than the 3 to 5 fps of motion resolution available to most humans. That is, if a spatial feature in the display of an animation only occurs in a single frame, it is likely not to be seen by the viewer unless it exists visually for 1/3 to 1/2 of a second. Conversely, animations begin to appear jerky at less than 5 fps. The proper specification for an animated query will take some thought with respect to the quality of the resulting visualization. An animated query language must be flexible enough to allow these factors to be taken into account during the query.

Other factors affecting query specification that impact display time include the granularity of stored temporal attributes and the combined seek rate of the mass storage and the data management system. For queries that are dependent on lookups of stored data, the ability to construct and redisplay frames in real time may be inhibited by slow database query performance or innately slow disk access operations. In such cases, it may be desirable to send the results of queries to a frame file (such as an MPEG format file) for post-construction replay. The components of display time then are: the desired duration (in real time) of the returned animation, the beginning and ending times bounding the temporal query, the computed or retrieved granularity of the data (or model) being queried, and the mapping of the returning granularity to the display granularity (optimized for visualization). At this point one must wonder if an animated query language can incorporate all of this without becoming too cumbersome.

Temporal Queries

The query language used by most relational systems is the Standard Query Language (SQL). The Standard Query Language (SQL) can be used to query virtually any relational database. While limiting temporal queries to very discrete, usually coarse- grained snapshots, the query language provides an adequate grammar and user-friendly syntax. On the object base side, the emerging Object Query Language (OQL) has yet to become as standardized among object bases as has SQL in the relational realm, but it is a start and utilizes a SQL grammar and syntax. These query languages are intended only to support static visualizations. Neither grammars allow for animated queries primarily due to the lack of a provision for display time.

Other lesser known query languages have been proposed that deal primarily with the issues of valid vs. transaction time and even include new spatio-temporal data models (Worboys, 1994; Snodgrass, 1992; Hazelton, 1992). Working back, there has been a noticeable literature addressing the need for animation in information systems (Rex, 1993, 1994; Okazaki, 1993; Flewelling, 1992), but so far there has not been a literature unifying these concepts. While it is beyond the scope of this paper to tie a specific spatio-temporal data model to the concept of an animated query language, the AQL prototype offered here should be general enough to support most of the more popular models based on simplical complexes and, to some, degree the current 2.5D relational model.

There are several assumptions taken here which must be specified. The first assumption is that the underlying system is object oriented. So that we need not get wrapped up in a discussion of topology, it is assumed that there is an underlying complete topological algebra and that topological objects can be referred to by name in the data store. Temporal attributes can have their values determined at runtime via class methods. Issues related to multidimensional time are not explicitly addressed, although one may easily imagine an extension to AQL in support of 2D or other time. It is assumed that time is linear and one-dimensional. Finally, it is assumed that every spatial object inherits a virtual time member function that is overridden by a local class specific method at runtime.

Any spatio-temporal query must be comprised of three primary components. First, the query must contain spatial bounds. An area or volume must be specified as being the object of interest. The query must also contain temporal bounds. If a query is temporally static, an event time must be specified. If the query is over some temporal interval, beginning and ending times must be specified. The query will most likely contain (or be able to handle) qualifying attributes of the object of interest. These primary query components, bounding space, bounding time and qualifying attributes will be augmented with one additional component for animated queries: display time.

Animation query, in this presentation, is not intended to replace traditional queries which do not result in visualizations. It is intended to assist in analysis through simplifying data organization needed to support visualizations. However, data may be reduced by assigning the results of an animated query to another object instead of generating an animation. In most cases, a standard SQL or OQL query would probably suffice. The important part of the AQL is the provision for display time and its subsequent ability to map returning query results to viewable animations.

As with any query language the prototype Animation Query Language (AQL) has a grammar based on lexicon and syntax. The current implementation of AQL consists of only a few keywords, but, used in the proper syntax, these can provide support for relatively sophisticated queries. The current keywords and operators are:

> **ANIMATE, BEGIN, DISPLAY, END, EXISTS, FOR, IN, INCREMENT, NULL, SPACE, TO, =, <, >, <=, >=**

and are used in the following syntax:

> **"ANIMATE** object **FOR** object attribute operator comparator(object attribute) **IN** object **SPACE BEGIN** time value **INCREMENT** time value **END** time value **DISPLAY** animation duration (in seconds) **TO** object, screen, or file name."

which provides for specifying an object of interest, the bounding space and time, the qualifying object attributes, display granularity, animation length, and an object or file name for storing the animated query. The INCREMENT keyword is used by the object of interest's time member function to either seed the calculation of time-dependent object attributes or retrieve stored values that match (or are close to) the time increment. Additionally, it effects the granularity of the returned animation. Assuming that the time member function takes the DISPLAY value as an argument, the resulting frame rate would be the total number of increments iterated through by the time member function divided by the DISPLAY time value. By adjusting these two values, optimal animations can be achieved. The EXISTS and NULL keywords are used as comparators for testing the values of object attributes. TO is optional.

As an example of the application of an AQL query in 2D+time, suppose we are analyzing a database containing the output from a hydrologic modeling package. We are interested in the temporal patterns of soil moisture distributions during spring snowmelt in a local area of a larger watershed model. Data has been generated and stored in our object base in snapshot style as a time value for the sim_time object attribute. A sample query might be:

> "**ANIMATE** soil_moisture **FOR** sim_time **= EXISTS IN** interest_area **SPACE BEGIN** 03-01-90:00.00 **END** 6-18-90:18.00 **INCREMENT** :12.00 **DISPLAY**

10."

where "soil_moisture" is the object of interest, "sim_time" is the object of interest's qualifying attribute value, simulation time, EXISTS indicates that any existing soil moisture map qualifies, "interest_area" ties the extents of the viewport to the extents of the watershed subregion, BEGIN and END specify a search for data generated for between midnight March 1 and 6:00 PM June 18, inclusive. An INCREMENT of 12.00 requests a check of the object base at twelve-hour resolution, while the keyword DISPLAY specifies the generation of a 10 second animation of the query results. The returned query would be an animation showing 2D soil moisture distribution changes in the region of interest during the specified temporal window. The effective frame rate would be 220 frames (110 days X 2 frames/day) divided by 10 seconds, or 22 frames per second. This might be somewhat fast, so the query could be respecified with a display time of 44, which would give an effective frame rate of 5 fps, which would be easier to view.

Note that the time member function can be a submodel that calculates attribute values as needed if they do not currently exist in the object base. In the preceding example, suppose that an examination of the resulting data animation showed very rapid changes in the soil moisture distributions occurring over a short period of time. The query could be respecified over the duration of the phenomenon in question at a finer temporal granularity, say on an hourly basis. If the data did not exist in the database, the query engine could initialize and run the appropriate submodel (using existing data to set the model boundary conditions) to generate the requested animation.

As a further example of this capability, consider that we wish to analyze the temporal pattern of prairie grass growth which has been previously modeled as a function of precipitation and temperature. The grass class time member function takes the increment value, beginning and ending times as input arguments and returns a grass height for all grass locations over the specified time. Every grass object has a location as an object attribute. The AQL query might look like the following:

"**ANIMATE** grass **FOR** grass_height > 0.5 **IN** gzone_5 **SPACE BEGIN** 2-1-94:00.0 **END** 9-30-94:00.0 **INCREMENT** weekly **DISPLAY** 60."

In this example, the INCREMENT value "weekly" would have been predefined in the grass class time member function. This query would return an animation of grass height in gzone_5 (with preassigned colors for heights) with an approximate frame rate of (24 weeks / 60 seconds) 0.4 fps. This frame rate is like watching grass grow! A more interesting animation would result in shortening the animation by reducing the DISPLAY time rather than making the temporal granularity of the INCREMENT value finer - grass only grows so fast.

Thus, domain-specific submodels can be incorporated directly into the class structure and work hand-in-hand with animated queries. As shown, AQL is also effective when applied to snapshot temporal data. One potential drawback to the use of integrated temporal submodels is that if the user specifies an INCREMENT granularity that is too fine, such as specifying the grass growth calculation be made on minute intervals instead of intervals of a week, unnecessary calculation will take place with little or no improvement in the animation resulting from the query. The animated query will return a static image if the BEGIN and

END time values are the same. This will be true for either a snapshot database search or a member function submodel - either will be passed just one time interval. The DISPLAY value will permit the single returned frame to be viewed for as long as specified.

Conclusions

Advances in computer hardware in the areas of multiprocessing and graphics and the maturation of the object paradigm in software can combine to make compute-intensive submodels integrable to visualization information systems. This is consistent with the current move toward higher dimensional computer interfaces such as in virtual reality, 3D displays and multimedia. Some form of animated query will be expected as computer technology advances further. Although the AQL prototype presented here is a step down the path, it obviates the need for thought concerning the use of animation in spatio- temporal analysis and system design.

Some of the issues needing further study are human/computer interface problems regarding interactive object selection in 3D, command line queries vs. forms-based queries vs. visual queries (incorporation of a visual animated query language), complex queries, extensions to the language to give more visual control such as object colors and scene perspective, freeze frames and single-time-stepping through queries. Other issues deserving research include the use of animated queries in distributed applications such as World Wide Web, CORBA and other wide-area applications. Distributed processing begs the question of standards. Can a standard be developed for AQL that will ensure compatibility between languages and development environments such as HTML, VRML, C++, and Smalltalk?

Given continued research and development of the Animation Query Language, spatio-temporal analysts should soon have a new methodology with which to ply their craft. The current general trend in the information system industry toward "object-ness" is encouraging, and analysts are beginning to abstract problems in object terms as well. The linking of such products as AVS, SPYGLASS and MATHWORKS to GIS show that more advanced visualizations as well as other analytic tools are being sought by analysts. Hopefully, the vendor community will begin to move toward providing some form of AQL in their products. Because there are currently very few object-based visual information systems in use, it may take some time before AQL shows up generically.

References:

Buttenfield B., 1993, Proactive Graphics and GIS: Prototype Tools for Query, Modeling and Display, Proceedings of AutoCarto 11.

Chrisman, N., 1993, Beyond Spatio-temporal Data Models: A Model of GIS as a Technology Embedded in Historical Context, Proceedings of AutoCarto 11.

Falconer K., 1990, Fractal Geometry: Mathematical Foundations and Applications, John Wiley and Sons.

Flewelling D., Frank A. and Egenhofer M., 1992, Constructing Geologic Cross Sections with a Chronology of Geographic Events, Proceedings of the Fifth Annual Symposium on Spatial Data Handling, IGU Commission on GIS.

Kemp Z., and Thearle R., 1992, Modelling Relationships in Spatial Databases, Proceedings of the Fifth Annual Symposium on Spatial Data Handling, IGU Commission on GIS.

Langran G., 1991, Time in Geographic Information Systems (Technical Issues in Geographic Information Systems), Taylor and Francis.

Okazaki D., 1993, Visualizing Geographic Data Through Animation, Proceedings of AutoCarto 11.

Pigot, S. and Hazelton, B., 1992, The Fundamentals of a Topologic Model for a Four-Dimensional GIS, Proceedings of the Fifth Annual Symposium on Spatial Data Handling, IGU Commission on GIS.

Rex B., 1994, 4DIVAS: Four Dimensional Information Visualization and Analysis System, Proceedings of ACM GVIS '94, Richland, Wa., pp. 59-62.

Rex B, 1993, Object Parallel Spatio-Temporal Analysis and Modeling System, doctoral dissertation, Portland State University, Portland, Oregon.

Slocum T., Davis J., and Egbert S., 1993, Developing Software for Exploring Temporal Spatial Data, Proceedings of LIS/GIS '93, Minneapolis, Minnesota, 1993.

Snodgrass R., 1992, Temporal Databases, from "Theories and Methods of Spatio- Temporal Reasoning in Geographic Space", Frank A., Campari I., and Formentini U., eds., Springer-Verlag.

Worboys M., 1994, A Unified Model for Spatial and Temporal Information, Computer Journal, Vol. 37, No. 1.

May Yuan, Department of Geography, The University of Oklahoma

Temporal GIS and Spatio-Temporal Modeling

Abstract

This paper investigates the development of temporal GIS and its applicability to support spatio-temporal modeling. Many GIS data models have been proposed to incorporate temporal information into spatial databases. Their general frameworks, with littl e considerations on data needs for spatio-temporal modeling, use a set of geometry-based spatial objects to represent reality. Thematic characteristics are represented as attributes of spatial objects. Temporal information is either associated with time-s tamped individual layers, such as the Snapshot Model (Armstrong, 1988), or individual spatial objects, such as the Space-Time Composite Model (Langran and Chrisman, 1988). The snapshot approach usually results in significant data redundancy. The space-tim e composite approach requires re-construction of thematic and temporal attribute tables whenever operations involve any changes in spatial objects (shape, size, or configuration). Consequently, geographic entities tend to be decomposed into fragments of s patial objects. For example, a wildfire event can be represented by a set of polygons with descriptions of burn severity and burn time. Problems appear, however, in representing phenomena like front lines, re-ignition, and spotting. The lack of direct map pings from GIS data to model input hampers GIS capabilities of spatio-temporal analysis, such as calculating periodicity, rate of movement, and process. Further attempts have been made to provide such direct mappings by event-based (Peuquet and Duan, 1995) or object-oriented data models (Worboys, 1992, Raper and Livingstone, 1995). This paper aims to (1) examine these typical temporal GIS data models, (2) discuss their applicability to facilitate spatio-temporal analytical modeling, (3) provide examples o f spatio-temporal analytical models and phenomena difficult to be handled in these proposed temporal GIS data models, and (4) synthesize important data concepts in spatio-temporal analytical models to be included in temporal GIS.

I. Introduction

A temporal GIS aims to process, manage, and analyze spatio-temporal data. However, the capabilities of any information system largely rely on the design of its data models. Data models present the conceptual core of an information system; they define data object types, relationships, operations, and rules to maintain database integrity (Date, 1995). A rigorous data model must anticipate spatiotemporal queries and analytical methods to be performed in the temporal GIS. Information about temporal constr ucts must be represented by data objects defined in data models to be stored or retrieved for analysis in a GIS. If a temporal GIS does not have a good data model, its support for temporal queries and temporal analysis of phenomena will be ineffectual.

Conventional GIS data models emphasize static representations of reality. Geographic information for a given area is decomposed into a set of single-theme layers as regular (raster) or irregular (vector) tessellation models (Frank and Mark, 1991). The se layers constrain GIS capabilities to represent dynamic information, such as transitions and motion. Raster cells encode attribute values at every given location with no considerations of the spatial characteristics of the theme they represent. Geometri cally indexed vector objects, on the other hand, 'force a segmentation of the entities being represented into separate layers whenever they interact in time or space: adopting this representational method forces compromises on most environmental modeling' (Raper and Livingstone, 1995, pp. 359). GIS needs a complete and rigorous framework for geographical data modeling

(Goodchild, 1992) to overcome the difficulty in handling geographic complexity, scale differences, generalization, and accuracy (Burrough a nd Frank, 1995). The lack of data representation schemata to integrate GIS data with models for spatiotemporal processes appears to be a major shortcoming in current GIS.

The paper discusses the trend of modeling temporal data in GIS. Key temporal GIS data models will be elaborated on the development of modeling spatiotemporal data. The discussion that follows provides difficult cases to be represented by the current t emporal GIS data models. To improve the capabilities of temporal GIS, the paper suggests important constructs in the representation of spatiotemporal phenomena.

II. The Trend of Temporal Data Modeling in GIS

The development of temporal data modeling in GIS parallels to the progress of temporal data modeling in the computer science (CS). The incorporation of temporal components has been implemented with the relational model and then with the object-oriente d data models in CS. The trend of temporal data modeling in GIS is moving from time-stamping layers (similar to relational tables) to time-stamping events or processes (similar to objects). Data semantics is a key issue in data modeling and the fundamenta l idea is to raise the level of abstraction in the transition from layers/tables to events/objects.

II.1. Representing spatiotemporal information by time-stamping spatial objects

In the last decade considerable research effort has been directed to temporal databases and temporal query languages (Tansel et al., 1993). Computer scientists have proposed methods to incorporate temporal information into a relational database by tim e-stamping a relation (a table, Gadia and Vaishnav, 1985), individual tuples (ungrouped relations, Snodgrass and Ahn, 1985), or individual cells (grouped relations, Gadia and Yeung, 1988). In GIS, Langran and Chrisman (1988) and Langran (1988) explore the idea of temporal GIS to outline a framework for conceptual design and implementation of incorporating temporal information in GIS. As Snodgrass (1992) summarizes the progress on the support for time-varying information in database management systems, Lan gran (1993) provides a comprehensive and compelling synthesis of temporal research in GIS. Parallel to the three relational database approaches, temporal information has been incorporated into GIS spatial data models by time-stamping layers (the snapshot models, Armstrong, 1988), attributes (space-time composites, Langran and Chrisman, 1988), and spatial objects (spatiotemporal objects, Worboys, 1992).

In the snapshot model, every layer is a collection of temporally homogeneous units of one theme (Figure 1). It shows the states of a geographic distribution at different times without explicit temporal relations among layers. Time intervals between an y two layers may vary and there is no implication for whether changes occur within the time lag of any two layers. The Temporal Map Sets (TMS, Beller et al., 1991) model can be seen as extensions of the snapshot model. The design of TMS purports to model geographic events in a defined area (Figure 2). Events are defined as binary TMSs, specifying whether each cell is in or out of the event. These snapshot approaches always result in a large amount of data duplication with unchanged properties in space and time. The major drawback is data redundancy and the risk of data inconsistency.

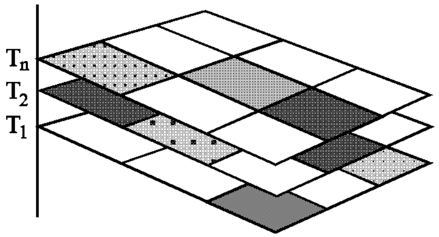


Fig.1: An example of the snapshot model (Armstrong, 1988)

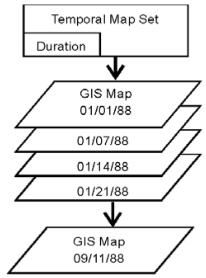


Fig.2: An example of a TMS (Beller et al., 1991)

The model Space-Time Composites (STC) represents the world as a set of spatially homogenous and temporally uniform objects in a 2D space (a layer, Figure 3). Every space-time composite has its unique temporal course of changes in attributes. Apparentl y, space-time composites can be derived by temporal overlays of time-stamped layers (snapshots). A space-time composite conceptually describes the change of a spatial object through a period of time. Attribute changes are recorded at discrete times, altho ugh its temporal resolution is not necessarily accurate. The STC model is able to record temporality within the largest common units of attribute, space, and time (i.e. change in situ), but it fails to capture temporality among attributes across space (i. e. motion or movement). In addition, updating a database of STC requires reconstruction of STC units. Consequently, geometrical and topological relationships among STC units change and the whole database, both spatial objects and attribute tables, needs t o be re-organized.

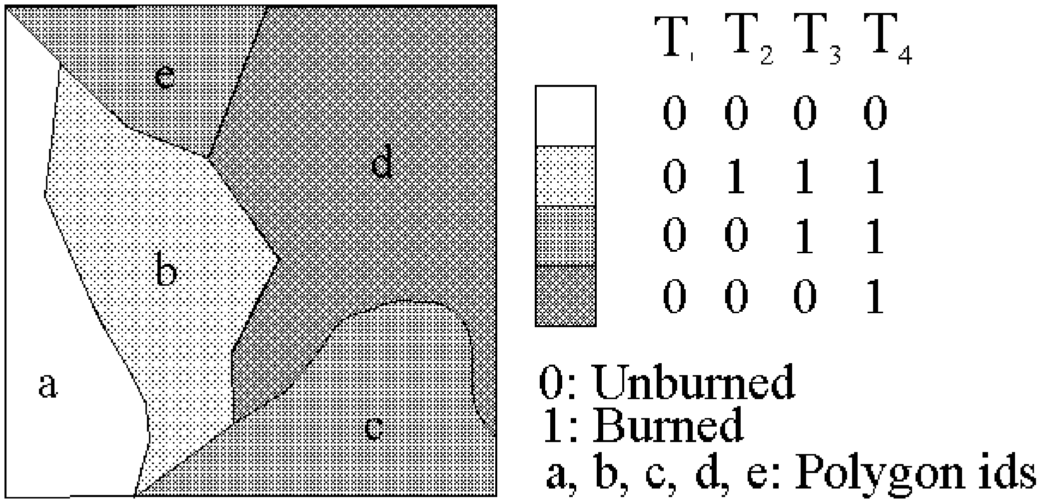
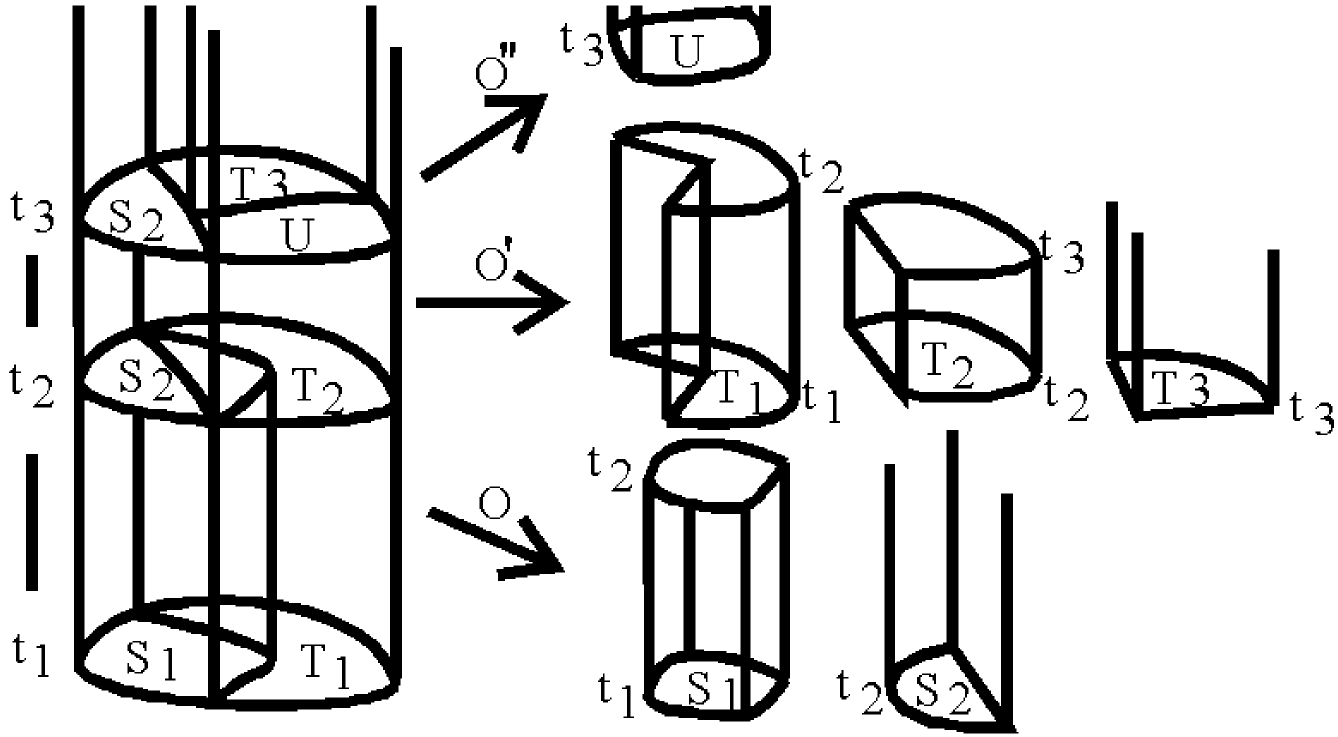


Fig.3: An example of an STC layer for burns (Modified from Langran and Chrisman, 1988)

The spatiotemporal object model (ST-Object model) represents the world as a set of discrete objects consisting of spatiotemporal atoms by incorporating a temporal dimension orthogonal to the 2D space (Figure 4). Spatiotemporal atoms are the largest ho mogeneous units in which certain properties hold in both space and time. A spatiotemporal object can possess changes in both space and time, although there is no change occurring within each of its spatiotemporal atoms. Therefore, the ST Object model is a ble to record changes in attributes of a ST-object in both spatial and temporal dimensions, together or separately, by projecting its ST-atoms to the spatial and/or temporal space. However, gradual changes in space through time are unable to be represente d in the ST Object model since its ST-atoms are discrete. Though the ST Object model is similar to the snapshot model and STC model, it only represents sudden changes upon an independent, discrete, and linear time structure. None of them are able to portr ay the concepts about transition, process, or motion.



ST-objects modeling regional change

Decomposition of ST-objects into ST-atoms

Fig.5: An example of a spatiotemporal object model with spatiotemporal atoms (Worboys, 1992)

II.2. Representing spatiotemporal information by events or processes

Computer scientists have developed temporal data models based on the concepts of time sequences (STC: temporal sequence collection, Segev and Shoshani, 1993), and time objects (OODAPLEX, Wuu and Dayal, 1992). These models are comparable to the recent studies of temporal information modeling by events or processes, such as the event-based spatio-temporal data model (ESTDM, Peuquet and Duan, 1995), and the geomorphologic spatial model (OOgeomorph, Raper and Livingstone, 1995).

Peuquet and Duan (1995) proposed a new raster-based data model (ESTDM) to organize spatiotemporal information about locational changes (Figure 5). Both ESTDM and TMS models group time-stamped layers to show temporal observations of a single event in a temporal sequence. However, ESTDM outperforms TMS in terms of data efficiency and support for analysis of temporal patterns and relationships, since the ESTDM stores 'changes' in relation to a previous state rather than a snapshot of an instance. A heade r file contains information about its thematic domain, pointer to a base map, and pointers to the first and last event lists. The base map shows an initial snapshot of a single theme of interest in a geographic area. An event-based series (EST series) con sists of the spatiotemporal dynamics of the thematic domain in that geographic area. Every event is time-stamped and associated with a list of event components to indicate where changes have occurred. An event component shows changes to a pre-defined loca tion (a raster cell) at a particular point in time. The ESTDM has shown its capabilities and efficiency to support both spatial and temporal queries. However, the adoption of the ESTDM model to a vector-based system requires a substantial redesign of even t components. Historical or transitional information of an entity or a process will be fragmented if changes occur to spatial objects or their topology. Mechanisms are needed to allow event components to keep track of their pre-defined entities and locati ons.

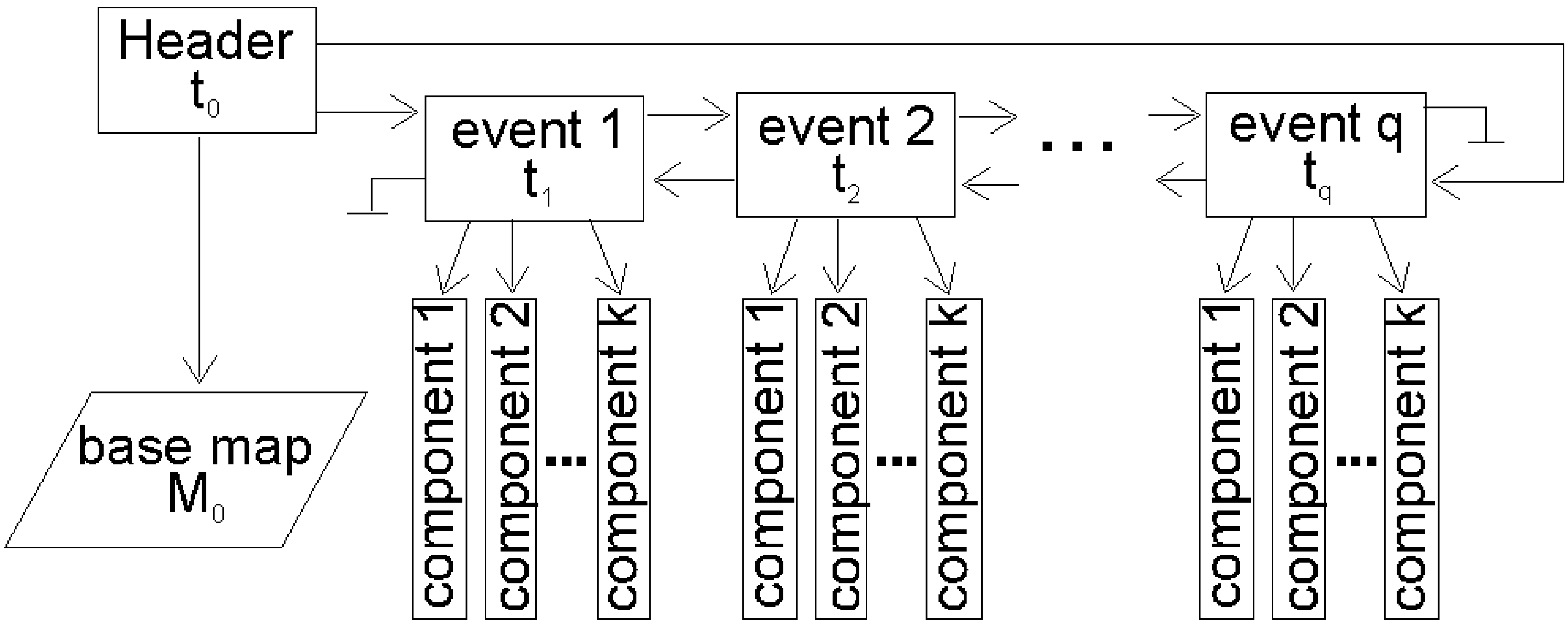


Fig.5: Primary elements and the pointer structure of an ESTDM (Peuquet and Duan, 1995).

While the ESTDM stores changes of a single theme at pre-defined locations, the design of OOgeomorph attempts to incorporate geomorphologic processes and theories with classes in an object-oriented representation (Figure 6, Raper and Livingstone, 1995) . A geomorph_info module models geomorphologic data from a geomorphologic spatial database to the data representation to be used in a geomorph_system module. A geomorph_system module represents the dynamics of a geomorphologic system, such as a coastal sy stem or a fluvial system, and includes associated geomorphologic theories. A set of CASE tools, mform, mprocess, and mmaterial, associate attribute data about geomorphologic forms, processes, or materials. Data input from the geomorph_info are used to ini tiate geomorphologic objects based on structures defined in classes of form, process, or material. As such, every geomorphologic phenomenon is represented by a set of form, process, and material objects, and every of these objects is, in turn, represented by a set of attribute objects. Three-dimensional location (x, y, z) and one-dimensional time (t) are referenced to objects in attribute classes (att). This approach is similar to Worboy's space-time object model of space-time objects and space-time atoms , but OOgeomorph stresses the importance of a physical system and processes within the system. Space-time objects and atoms are formed by their spatiotemporal associations, but objects in the OOgeomorph are linked by their relationships defined in a geomo rphologic system. While OOgeomorph can handle point-based locational information well, it has difficulty in manipulating area data and topological relationships.

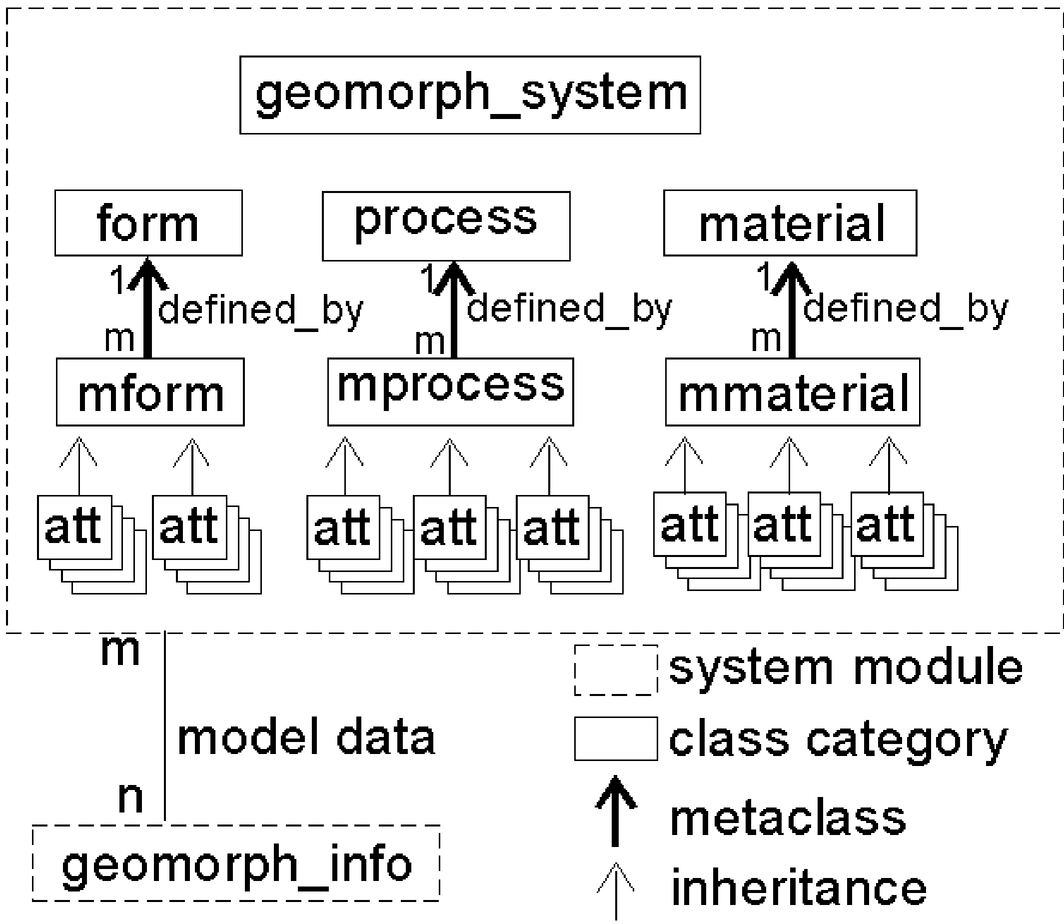


Fig.6: The design of OOgeomorph: an object-oriented data system (Raper and Livingstone, 1995).

However, the real challenge in GIS temporal modeling is to maintain spatial objects' identities throughout the evolution in geometrical properties and topological relationships. This is unimportant for aspatial databases or raster/point-based GIS data bases. Raster or point-based spatial objects are stationary in space and temporal information can be associated with spatial object identifiers and therefore can be handled mainly as an aspatial database. Line- or polygon-based GIS are very likely involve d in changes to geometry and topology of spatial objects. Domain-oriented data models are proposed to manage complex changes of spatial objects and maintain object identities.

The three domain model is developed by analyzing spatiotemporal information needs for wildfire studies and representational requirements to facilitate these studies in a GIS environment. A wildfire information cycle shows the needs of four data schema ta. Snapshots represent states, fire entities represent processes, entity snapshots represent changes, and fire mosaics represent history. These representational schemata are designed to support spatiotemporal data for wildfire studies of fire forecasting , fire behavior, fire impacts, and fire history. A separation of semantical1, spatial, and temporal information is necessary in order to dynamically support all four schemata. The three domain model defines semantical, temporal, and spatial objects in thr ee separate domains. Time is modeled as an independent concept in the three domain model, instead of being an attribute of location as in the snapshot model or being an integral part of spatial entities as in the space-time composites and spatiotemporal o bjects. Geographic concepts and entities are represented by dynamically linking the three types of objects from a layer or object perspective (Figure 7). The data model is able to represent reality from locational-centered, entity-centered, and time-cente red perspectives with six basic types of changes in geographic information: attribute changes, static spatial distribution, static spatial changes, dynamic spatial changes, mutation of a process, and movement of an entity (Yuan, 1995). The major advantage of the three domain model is that there is no pre-defined data schemata; rather the model will dynamically link relevant objects from the three domains to represent a geographic entity or concept. Linkages among these objects can be numerical or fuzzy me mbership functions. These functions are useful to represent dynamic boundaries such as transitions of soil distributions, seasonal changes in lake boundaries, or diurnal changes in shorelines.

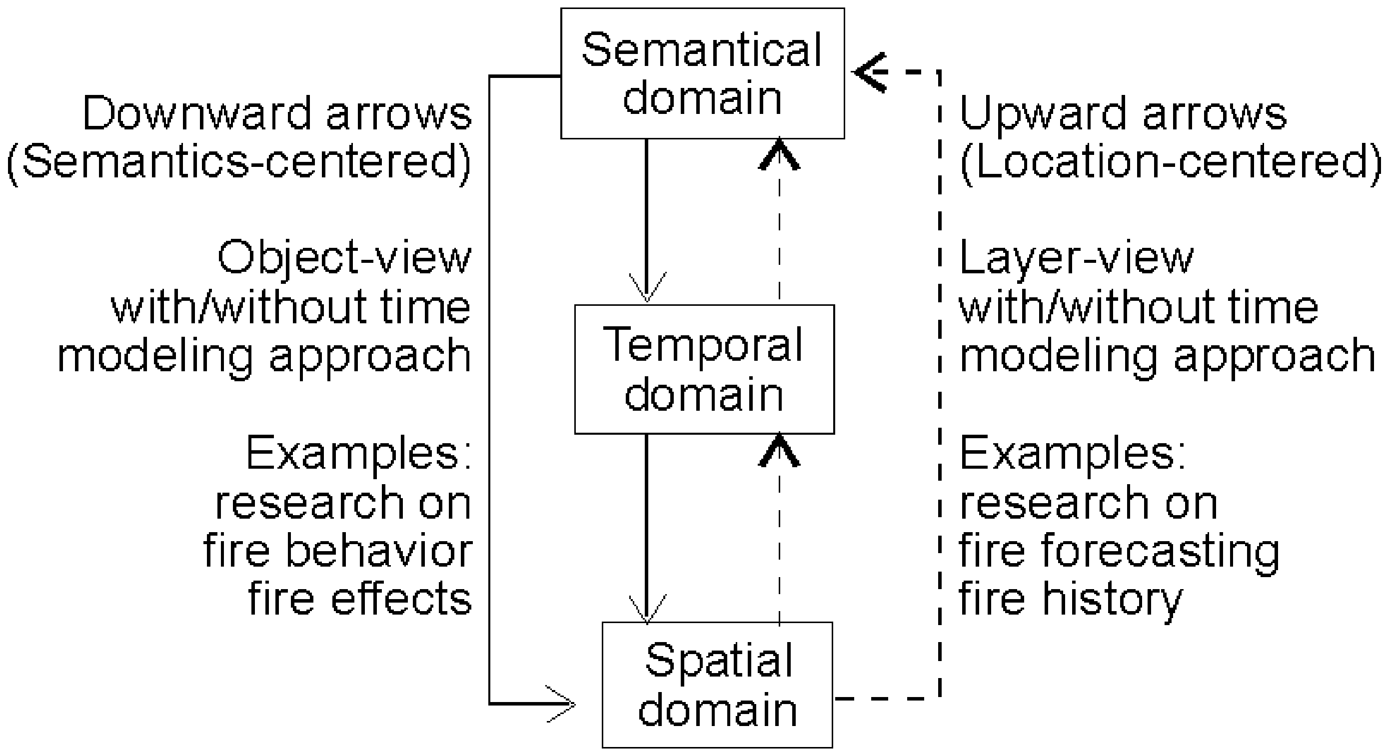


Fig. 7: An conceptual framework of the three domain model (Yuan, 1994b).

Smith et al. (1993) and Smith (1994) also take a domain approach in the design of a modeling and database system (MDBS) to support high-level modeling of spatiotemporal phenomena. The MDBS consists of a conceptual domain (C-Domain) for abstract views of entities and transformation and a representation domain (R-Domain) for symbolic representation (Figure 8). Typical R-Domains include primitive domains (bools), purely spatial domains (polygons), non-spatial domains (rainfall), geographic domains (drain age basins), and temporal domains (hydrographs). These R-Domains can in fact be incorporated into the domains of semantics, space, and time. The two systems are compatible, and the three domain model will benefit by incorporating the theory of domains as well as the modeling and database language (MDBL) developed in the MDBS.

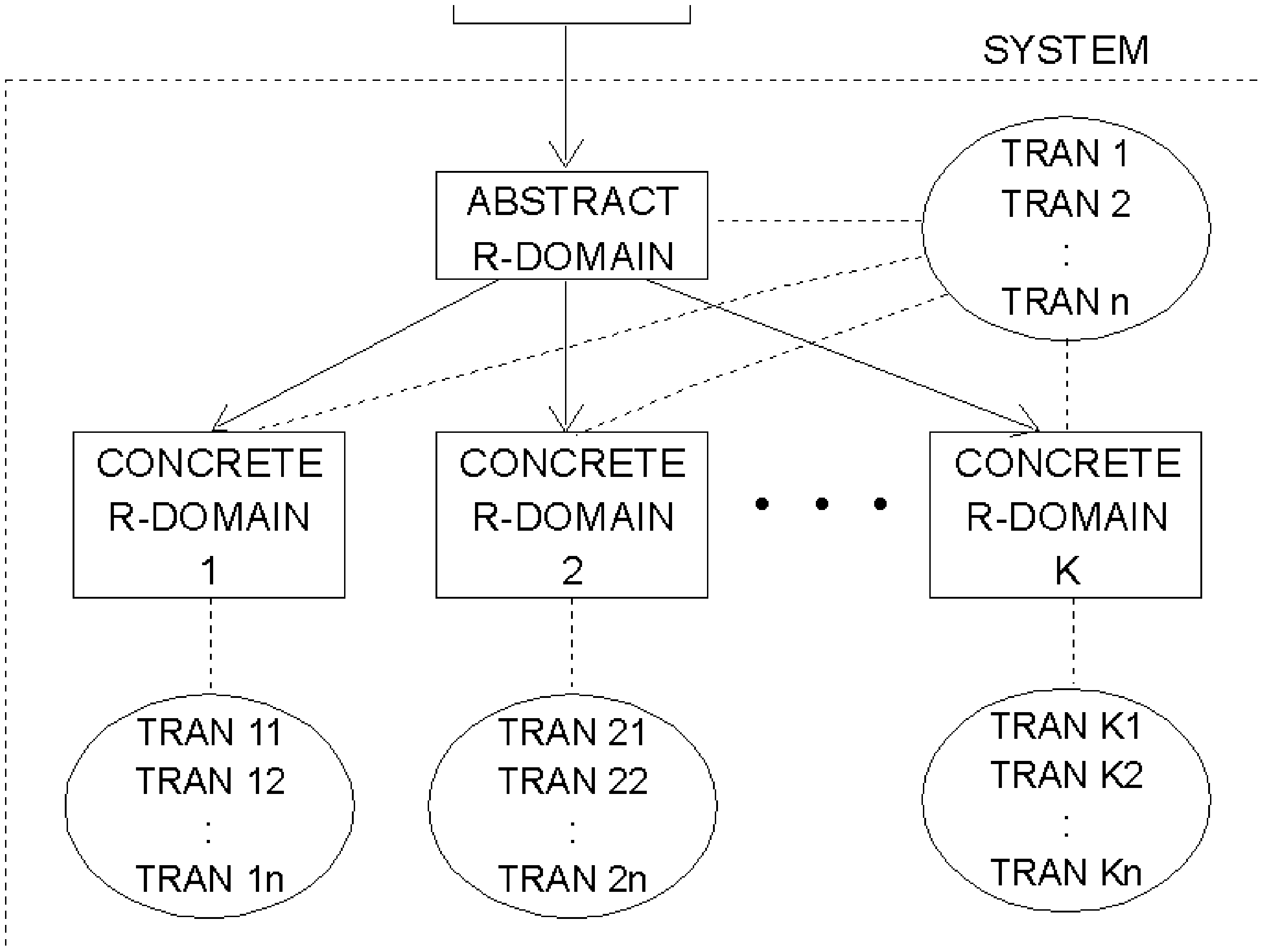


Fig.8: A C-domain, an abstract R-domain and its n concrete R-domains (Smith, 1994)

its n concrete R domains (Smith, 1994).

III. Examples of Difficult Cases for Temporal GIS

Changes occur to attributes of a phenomenon, environmental settings, behaviors of an event, or mechanisms of a processes. There are six major types of spatial and/or temporal changes in geographic information (Yuan, 1995):

I. For a given site where occurrences and duration of events or attributes may change from time to time, analysis is done by fixing location, controlling attribute, and measuring time.

II. For a given point in time where a certain phenomenon may change its characteristics from site to site, analysis is done by fixing time, controlling attribute, and measuring location.

III. For a given period of time where attributes may change from site to site through time, analysis is done by fixing time, controlling locations, and measuring attributes.

IV. For a given event where its characteristics or processes may change at sites through time, analysis is done by fixing attributes, controlling locations, and measuring time;

V. For a given area where attributes may change site to site and from time to time, analysis is done by fixing location, controlling time, and measuring attributes.

VI. For a given event where its location may change from time to time, analysis is done by fixing attributes, controlling time, and measuring locations.

Type I changes only involve variations in attributes over time; there is no variations in spatial properties. Therefore, this kind of data modeling and analysis can be totally handled in a semantical domain as historical transactions in a relational o r object-oriented database management system. Type II changes describe static spatial distribution of a geographic phenomenon, such as topography or air pressure. Techniques, such as contouring, choropleth mapping, and dasymetric mapping are often used to present such information, while current GIS store this kind of information in forms of vector or raster layers. Changes of Types III to VI alter geometry or topology of spatial and/or temporal properties.

III.1. Spatial Changes

Spatial changes refer to variations across space at a given time or in a period, in which comparisons are made between two or more sites according to data of the same vintage. Spatial changes can be classified as static (Type III) or transitional (Typ e IV). Static spatial changes concern variations of a geographic phenomenon at a snapshot, whereas transitional spatial changes compare states of an event or a process at different sites (Figure 9) . For example, we can compare vegetation successions at Site A and Site B to examine the ecological impact of air pollution. As another example, central and western India experience heavy rainfalls in El Nino years, while the northern India usually has deficit of precipitation during these periods. Type IV tra nsitional changes describe variations of spatial properties for a given event or process in a time series (Figure 10). Such changes can be shown by linkages from a set of temporal objects to a set of spatial objects. Three basic parameters in measuring sp atial changes are attribute, duration, and continuity of a phenomenon, event, or process. For example, we can compare transitional changes of landuse types at Site A with what has occurred at Site B from T1 to T2. Or, we can compare impact of the monsoon trough on rainfall processes in El Ni(o years at these two sites.

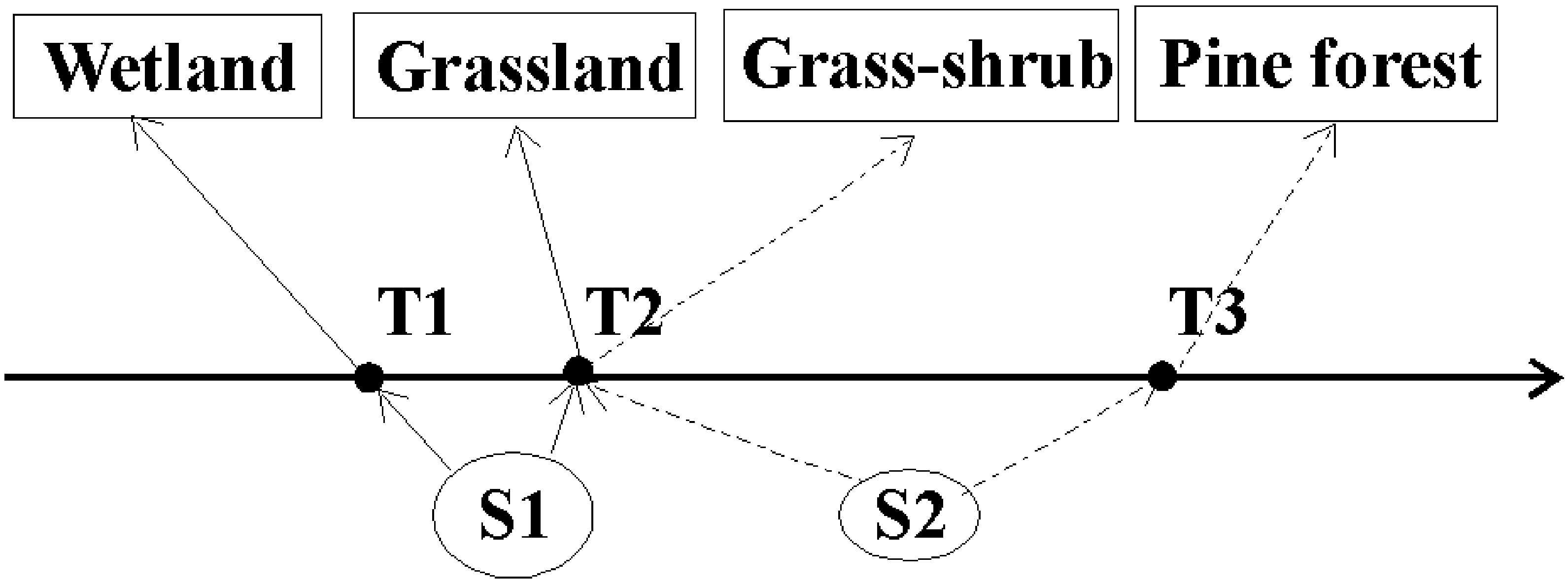


Figure 9: An example of static change (Type III Change).

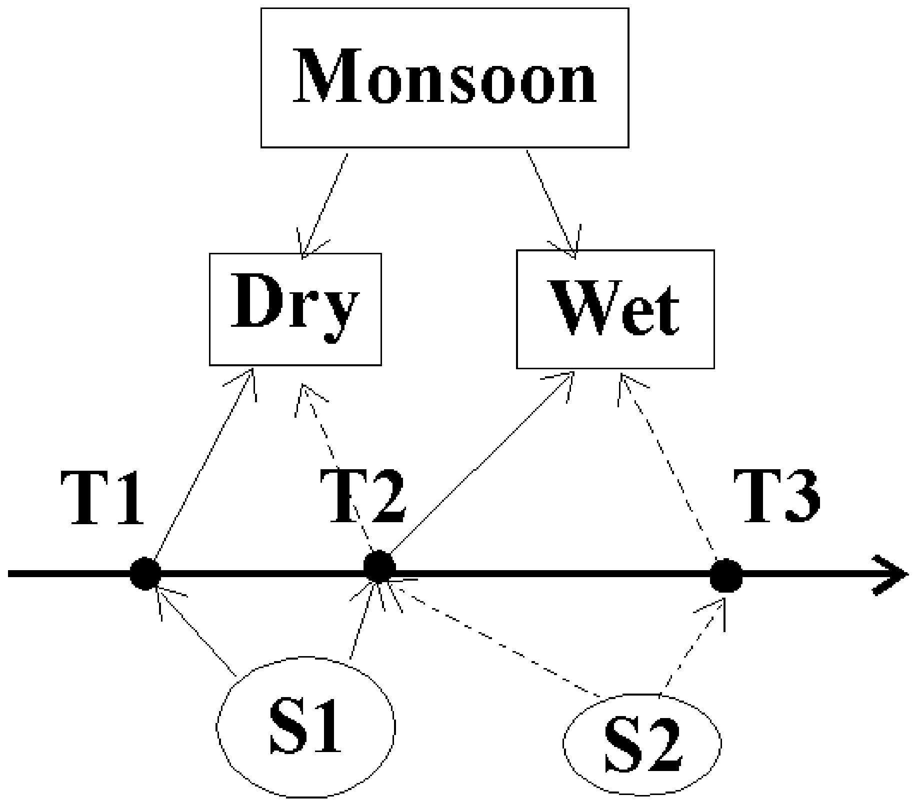


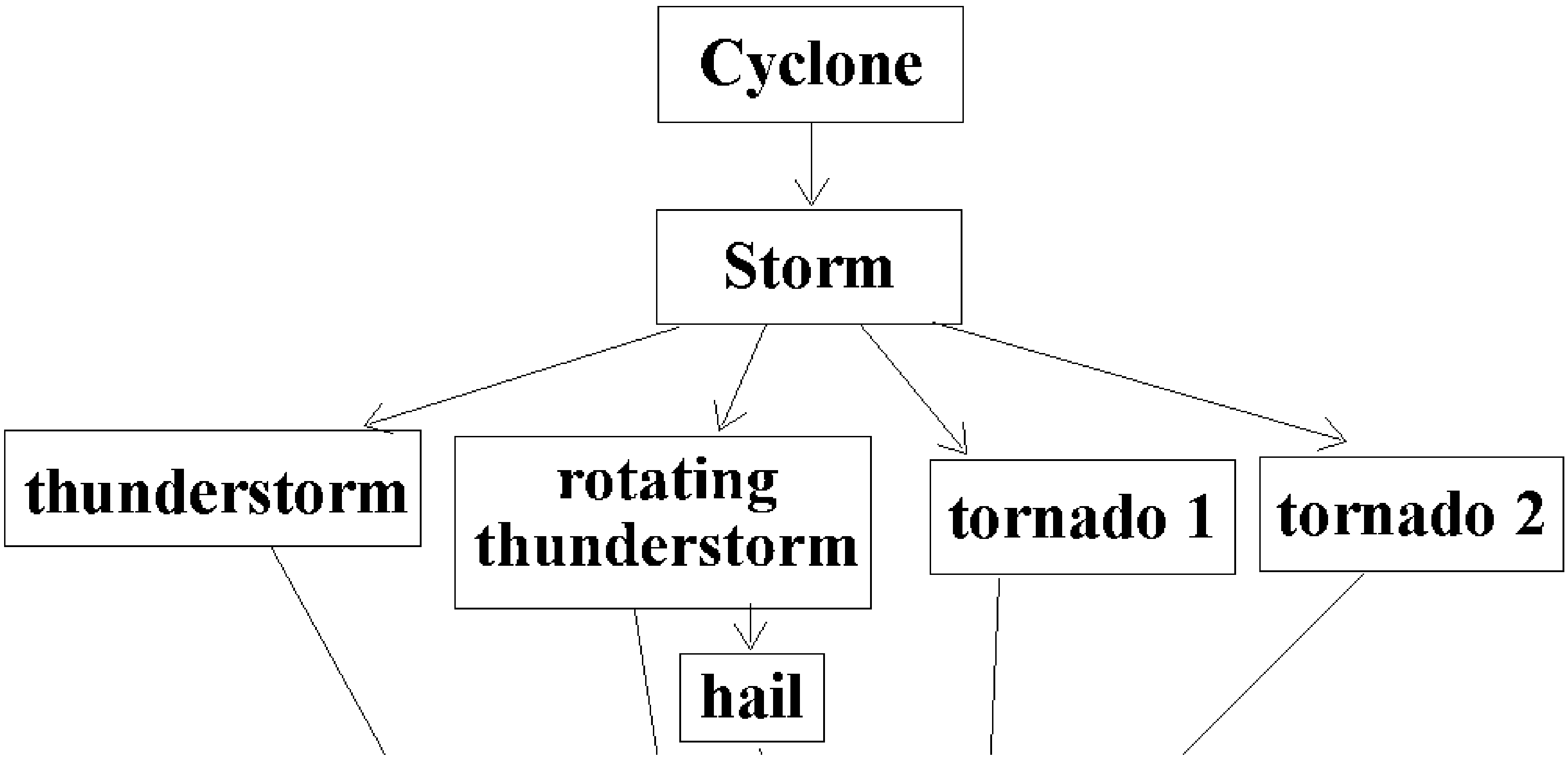
Figure 10: An example of transitional change (Type IV Change).

II.2. Temporal Changes

Temporal changes are changes occurring at different points or periods in time, and they are recognized by changes in spatial properties and/or locations from time to time. Two types of temporal changes are identified in this research: mutation (Type V) or movement (Type VI).

Mutation refers to changes occurring to the internal mechanisms of an event or a process, or to the interactions between events/processes and their environments. For example, a thunderstorm and a tornado may develop in a county and progress to other c ounties (Figure11). A practice of artificial rainfall in a region may mutate the patterns of precipitation and water resources in its surrounding area. Type V changes describe the mutation of a type of processes or events (semantical objects) by two sets of temporal objects; each of them is linked to a set of spatial objects. Comparisons are made between the two sets of spatial and temporal objects to show how a process mutates its attributes, temporal properties, and spatial characteristics in the two se ts of time series. Therefore, a comparison of frequency, period, and severity of an event in an area at different periods of time may suggest Type V temporal changes in the area.

Movement concerns the travel of an event or an entity from one place to another, and the event or entity may or may not involve changes in spatial properties other than location. Examples of movement include spread of wildfire, insect infestation, and animal movement (Figure12). The changes denote a single event or process and can be represented by linking a semantical object to a series of temporal objects and then to a set of spatial objects to show the movement of this event during the period const ituted of these temporal objects. Eight subtypes of temporal changes result from combinations of changes in attributes, morphology, and topology (Armstrong, 1988).



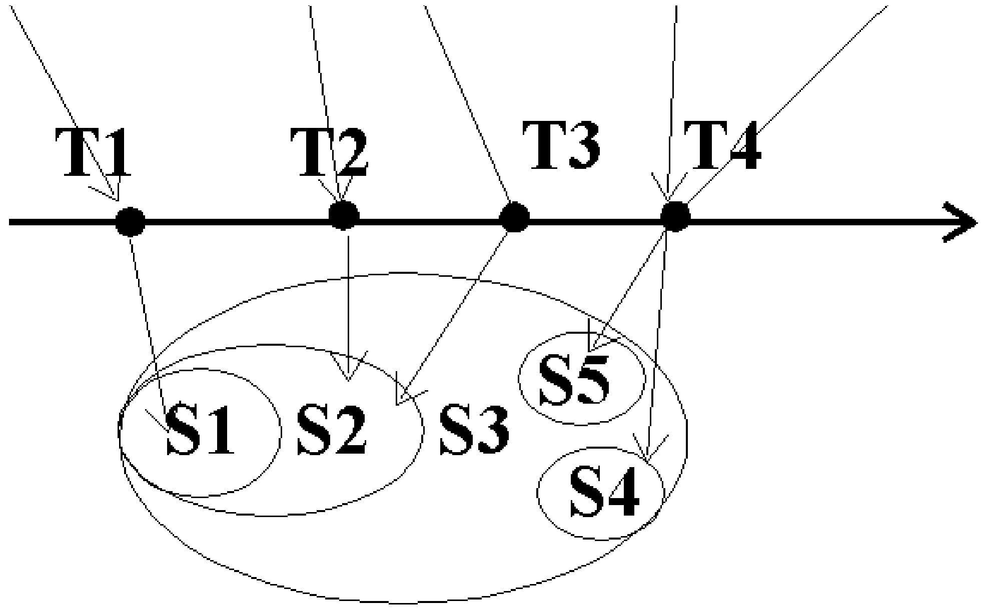


Fig. 11: An example of mutation (Type VI Change).

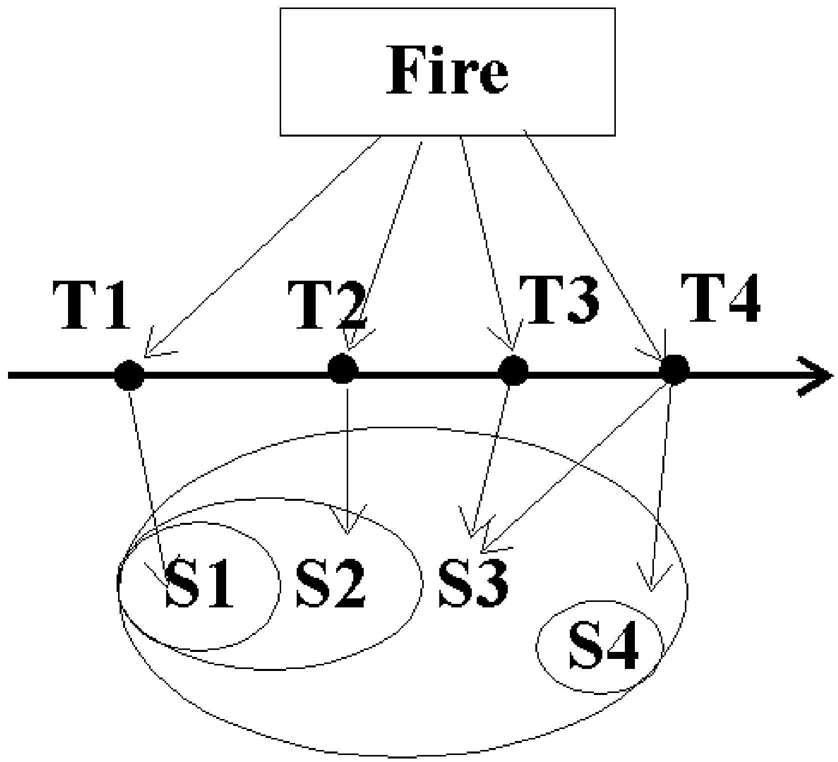


Figure 12: An example of movement (Type VI Change).

The incorporation of temporal elements into databases itself is a nontrivial task, and it becomes even more complicated when we try to model spatiotemporal data in GIS. The current temporal GIS data models lack of capabilities to represent Type III to Type VI changes. The following section discusses important spatiotemporal concepts to be considered in GIS data modeling to improve GIS's representativity by including important spatiotemporal concepts in its data models.

IV. Important Spatiotemporal Concepts to be Included in a temporal

GIS Data Model

Langran (1992) points out that '(p)recisely articulated information about what and where changes occurred within a geographic area is at the heart of a TGIS' (p. 419). In doing so, she suggests states, events, and evidence are the three principle ent ities of a temporal GIS. States describe the spatial distributions of a geographic phenomena. They can be represented by snapshots or space-time composites. Events cause changes of states for a given phenomena through time, such as floods or fires. Eviden ce describes how changes are discovered and measured. It provides basis for updating a state. Therefore, the combination of states, events, and evidence show what has been in an area, what has happened to the area, and how we know it has happened. The thr ee data types in fact constitute the key elements of temporal metadata in GIS.

States and changes of states are the main concern in temporal representation, although events and evidence are critical to ensure data quality and detect causal relationships. It is important to understand time when we try to record changes and associ ate aspatical and spatial objects with time. The representation of time is an important issue in the study of states and changes of states. Time can be conceptualized as instances (point-based) or intervals (Allen, 1983; Fresa, 1992). Time has multiple st ructures: linear time, cyclic time, parallel time, and branching time (Worboys, 1990). Time has multiple dimensions: valid time and transactional time, user-define time, and institutional time (Snodgrass and Ahn, 1986; Berrera et al., 1991). In order to r epresent multiple time dimensions, Worboys (1994) proposes a spatio-bitemporal model and operations to handle information about valid time and transaction time in GIS. Other important concepts about time are elaborated in Peuquet (1994) and Raper and Livi ngstone (1995).

In addition, temporal GIS data models need to incorporate the behaviors of natural phenomena as the examples given in section III. To date, the design of conventional GIS data models have neglected the physical processes of natural phenomena. The basi c concept in conventional GIS data models is "location." Basic GIS units are spatial objects (points, lines, polygons, and cells), and their static attributes. Spatial relationships or interactions are limited to being at the same location or in close pro ximity. In contrast, the basic concepts in modeling of natural processes are "system states," "mass and energy conservation," transformation and translocation," and "species and individuals' interaction and dynamics" (Fedra, 1993). GIS is, in its present state, unable to provide complete support for spatiotemporal modeling of natural processes. Temporal GIS should support both location-based analysis (Type III and Type IV Changes), such as landuse changes and ecological succession, and process-based chang es (Type V and Type VI Changes), for example fire spread and storm development. As such, locations, states, events, and processes should be handled as individual data constructs in a temporal GIS. Semantical analysis of natural phenomena on their characte ristics and behaviors is critical to determine a set of high-level spatiotemporal constructs to be modeled in a temporal GIS (Yuan, 1994b). Most of the current GIS data models, as described in section II, organize geographic information according to spati al objects. Attributes are associated with a cell, a point, a line, or a polygon. As a result, GIS does not represent Interstate 35 as a integral object, rather there are a set of independent line segments and they have an attribute value, Interstate 35. Consequently, information about changes to Interstate 35 can not be effectively represented in a temporal GIS.

It is important to understand the fundamental spatiotemporal concepts to be successful in modeling temporal data in GIS. Many GIS data models are developed from perspectives of structures in spatial data. However, a process-oriented or semantics appro aches may lead to a new direction in modeling geographic information. GIS will be more effective and precise by representing a fire, fire location, and how the fire spreads rather than burned areas at each point in time.

V. Concluding Remarks

The study of modeling temporal information in GIS started in the mid 1980s. Many data models have been proposed, and some of them have been implemented. The development of temporal data models in computer science has shown an influence on the trend of temporal modeling in GIS. However, GIS data modelers need to consider the evolution of spatial objects in addition to retroactive or postactive changes and all other issues to be considered in a non-spatial databases (Snodgrass, 1992). Geographic inform ation has three components: attributes, time, and space. While the three components can change and be analyzed independently (Berry, 1964; Sinton, 1978), the proposed temporal GIS data models, as reviewed in section II, cannot model all possible kinds of temporal information listed in section III. The Snapshot Model (Armstrong, 1988), Space-Time Model (Langran and Chrisman, 1988), Spatiotemporal Object Model (Worboys, 1992) all can represent states through time (Type I and Type II Changes). The Space-Tim e Model and Spatiotemporal Object Model can also represent state changes at locations (Type III Changes). None of them can precisely nor effectively model transitions, mutation, and movement of processes (Type IV, V, and VI Changes). While the Event-based spatiotemporal data model (ESTDM, Peuquet and Duan, 1995) and the OOgeomorph model (Raper and Livingstone, 1995) are able to show transitions, they have difficulty in handling mutation and movement. In addition, both of them are limited to raster or poin t-based GIS. Temporal GIS needs a top-down approach to modeling spatiotemporal information, because behaviors of natural phenomena need to be considered and should be considered prior to available GIS data formats and data structures in constructing tempo ral GIS representation. The process-oriented approach has been used in the development of the ESTDM, OOgeomorph, the three domain model (Yuan, 1994), and the modeling and database systems (MDBS, Smith et al., 1993) with a certain amount of success. Furthe r research still needs to incorporate spatiotemporal constructs from natural phenomena into the modeling of temporal information in GIS to fully represent Type III to Type VI changes in space and time.

References

Allen, J. F., 1983, Maintaining knowledge about temporal intervals. Communications of the ACM, 26(11): 832-843.

Armstrong, M. P., 1988, Temporality in spatial databases. Proceedings: GIS/LIS'88, 2:880-889.

Barrera, R., Frank, A., and Al-Taha, K., 1991, Temporal Relations in Geographic Information Systems. NCGIA Technical Report 91-4.

Berry, B. J. L., 1964, Approaches to regional analysis: A synthesis. Annals of the Association of American Geographers, 54(1):2-11.

Burrough, P. A. and Frank, A. U., 1995, Concepts and paradigms in spatial information: are current geographical information systems truly generic? International Journal of Geographical Information Systems, 9(2): 101-116.

Beller, A., Giblin, T., Le, K. V., Litz, S., Kittel, T., and Schimel, D., 1991, A temporal GIS prototype for global change research. Proceedings: GIS/LIS'91, 2:752-765.

Date, C. J., 1995, An Introduction to Database Systems, 6th edition (Reading: Addison-Wesley Publishing Company).

Fedra, K., 1993, GIS and environmental modeling. In Environmental Modeling with GIS, edited by . Goodchild , M. F, Parks, B. O., and Steyaert L. T. (New York: Oxford University Press).

Frank, A. U. and Mark, D. M., 1991, Languages issues for GIS. In Geographic Information Systems: Principles and Applications, edited by Maguire, D.

J., Goodchild, M. F., and Rhind, D. W., vol. 1 (Esses: Longman Scientific & Technical Inc.), pp. 147 -163.

Freksa, C., 1992, Temporal reasoning based on semi-intervals. Artificial Intelligence, 54:199-227.

Gadia, S. K. and J. H. Vaishnav. 1985. A query language for a homogeneous temporal database. In Proceedings of the ACM Symposium on Principles of Database Systems, pp. 51-56.

Gadia, S. K. and C. S. Yeung. 1988. A generalized model for a relational temporal database. In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 251-259.

Goodchild, M. F., 1992, Geographical information science. International Journal of Geographical Information Systems, 6(1):31-45.

Langran, G. and Chrisman, N, R., 1988, A framework for temporal geographic information. Cartographica, 25(3):1-14.

Langran, G., 1988, Temporal design tradeoffs. In proceedings of GIS/LIS'88, Volume 2 (Falls Church, VA: ACSM), pp. 890-899.

Langran, G., 1992, States, events, and evidence: the principle entities of a temporal GIS. In Proceedings: GIS/LIS'92, pp. 416-425.

Langran, G., 1993, Time in Geographic Information Systems (Bristol, PA: Taylor & Francis).

Peuquet, D. J., 1994, It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems. Annals of the Association of American Geographers, 84(3):441-462.

Peuquet, D. J. and Duan, N., 1995, An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. International Journal of Geographical Information Systems, 9(1): 7-24.

Rapper, J. and Livingstone, D., 1995, Development of a geomorphological spatial model using object-oriented design. International Journal of Geographical Information Systems, 9(4): 359-384.

Segev, A. and Shoshani, A., 1993, A temporal data model based on time sequences. In Tansel et al., eds. Temporal Databases: Theory, Design, and Implementation (Reading, MA: The Benjamin /Cummings Publishing Company, Inc.), pp. 248-270.

Smith, T. R., Su, J., Agrawal, D., El Abbadi, A., 1993, Database and modelling systems for the earth sciences. IEEE (6) (Special Issue on Scientific Databases).

Smith, T. R., 1994, On the integration of database systems and computational support for high-level modelling of spatio-temporal phenomena. In M. F. Worboys ed. Innovations in GIS (Bristol, PA: Taylor & Francis) pp. 11-24.

Sinton, D., 1978, The inherent structure of information as a constraint to analysis: mapped thematic data as a case study. Harvard Papers on GIS, Volume 7, edited by G. Dutton, Addison-Wesley Publishing Company, Inc., Reading, MA.

Snodgrass, R. and Ahn, I., 1985. A taxonomy of time in databases. In Proceedings of ACM SIGMOD International Conference on Management of Data.

pp. 236-264.

Snodgrass, R. and Ahn, I., 1986. Temporal databases. IEEE Computer, September, 1986, pp. 35-42.

Snodgrass, R. T., 1992. Temporal databases. In Frank, A. U. ed. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space (Berlin: Springer), pp. 22-64.

Sowa, J. F.. 1984. Conceptual Structures: Information Processing in Mind and Machine (Reading: Addison-Wesley Publishing Company, Inc.).

Steyaert, Louis T. and Michael F. Goodchild. 1994. Integrating geographic information systems and environmental simulation models: A status review. In William K. Michener, James W. Brunt, and Susan G. Stafford eds. Environmental Information Management and Analysis: Ecosystem to Global Scales (Bristol, PA: Taylor & Francis), pp. 333-356.

Tansel, A. U., Clifford, J., Gadia, S., Jajodia, S., Segev, A., Snodgrass, R., ed., 1993, Temporal Databases: Theory, Design, and Implementation (Reading, MA: The Benjamin /Cummings Publishing Company, Inc.).

Wuu, G. and Dayal, U., 1992, A uniform model for temporal object-oriented databases. In Proceedings of the International Conference on Data Engineering (Tempe, AZ), pp. 584-593.

Worboys, M. F. 1990, Reasoning about GIS Using Temporal and Dynamic Logics. NCGIA Technical Report 90-4.

Worboys, M. F., 1992, A model for spatio-temporal information. Proceedings: the 5th International Symposium on Spatial Data Handling, 2:602-611.

Worboys, M. F., 1994, A unified model for spatial and temporal information. The Computer Journal, 37(1):26-34.

Yuan, M., 1994a, Wildfire conceptual modeling for building GIS space-time models. Proceedings: GIS/LIS'94, pp. 860-869.

Yuan, M., 1994b, Representation of Wildfire in Geographic Information Systems. Unpublished Ph.D. dissertation. Department of Geography, State University of New York at Buffalo.

Yuan, M., 1995, Modeling semantical, temporal, and spatial information in geographic information systems. Under review for inclusion in Progress in Trans-Atlantic Geographic Information Research, First ESF-GISDATA and NSF-NCGIA Summer Institute in Geograp hic Information, a book to be published by Taylor & Francis, Bristol, PA.

1The word 'semantics' or 'semantical' is, in the context of data modeling, the meanings of objects or concepts in the physical or abstract world (Date, 1995). It is to emphasize the meanings of geographic features and their aspatial and atemporal relation ships.

Jonathan Baldwin, Peter Fisher, Joseph Wood and Mitchel Langford

Modelling Environmental Cognition of the View With GIS

Abstract

This paper explores some of the ways in which a cognitive appreciation of landscape may be matched to the GIS context of DEM analysis. Physiographical characteristics of landscape cognition can be modelled using the technology associated with viewshed analysis. Relief, depth of view, horizon characteristics and shape could all be measured using GIS functionality. Uncertainty in surface feature classification may be identified by examining scale dependencies in the landsape and storing the results as feature membership functions. It is suggested that cognitive criteria such as drama, mystery and coherence may have measurable surrogates by using the modelled view as a basis for their definition.

Introduction

Geographical Information System (GIS) functionality faces a major challenge. Currently most GIS operations are deterministic and precise, not allowing for flexibility concerning object size, spatial extent or functional outcome, although research has been undertaken on handling error and probability (Fisher,1995). Within GIS it remains difficult to represent a cognitive environment, where qualitative reasoning is an integral component of theoretical models.

There are many aspects of non-deterministic decision making within the sphere of spatial information. This paper will take the example of landscape within GIS, and explore some ways to convey a subjective analysis of the human perception to it. Current GIS functions used to address landscape tend to have their foundation based upon Digital Elevation Models (DEMs) or surrogates thereof (Bishop and Hulse, 1994). They appear to have been dominated by the hydrological modelling community relying on deterministic functions to isolate Boolean concepts such as the drainage network and the catchment area. Generation of the visible area has developed as part of this modelling functionality (Lange, 1994), and in the usual implementation of the operation shows this deterministic heritage (De Floriani and Puppo, 1994).

Visible area analysis is thought to be among the most widely used function in landscape planning with GIS (Davidson *et al*, 1993). Given the possible discrepancy between the objective result of visible area analysis in GIS and our subjective landscape perception, it is perhaps surprising that such an approach has been so broadly adopted. It is the contention of this paper that future guidance given to landscape planners and architects using GIS in relation to visual experiences needs to be enhanced by the introduction of cognitive elements within the digital environment. Further discussion will suggest reasons for this and how problems associated with the combination of such different data types may be addressed.

Significant research has been undertaken in the area of landscape cognition as it relates to personal experience of landscape (Daniel and Boster, 1976). The literature relating to landscape architecture and environmental psychology abounds with qualitative phrases such as interest, drama, mystery and quality (Kaplan and Kaplan, 1982; Preece, 1991) but it appears that the identification of the spatial extent of many classes of landscape feature (such as valley and hill) remains inconsistent between individual approaches (Shafer *et al*, 1969, 1973, 1974). No attempt known to the authors has been made to address this problem or model such cognitively defined elements within GIS.

What we can do already

Of foundational importance to the approaches discussed in this paper are the existing GIS functions used to assess the viewshed, distances, area and spatial distribution characteristics of a particular landscape (Miller *et al*, 1994). If two locations in a landscape enjoy an uninterrupted view of each other, then they are said to be in view from each other. When one location is specified as the viewing location, and the visibility of all other locations in the study area is analysed, the resulting map of the study area is known as the viewshed or the visible area.

The viewshed operation has received considerable attention with respect to its operational reliability (Fisher, 1991, 1992, 1993) and the optimisation of the line-of-sight operation (De Floriani *et al*, 1994; Lee, 1991, 1994). Significantly, however, this research has remained in the realm of deterministic modelling. Most assume that a location is either visible or it is not, although Fisher (1994, 1995) has explored a probabilistic approximation and its application and the use of fuzzy set membership functions has been proposed to determine a subjective version of the viewshed.

Researchers in environmental perception have concluded that personal experience of landscape can be classed into four general categories: *physiographical characteristics*, *the presence of specific physical features*, *cognitive variables and viewer interest* (Kliskey and Kearsley, 1993). However, the comparatively basic physical elements of relief, depth of view and the identification of specific features within a given visible area remain difficult to accurately describe in both subjective and digital environments. Cognitive variables such as drama, mystery and coherence (Kaplan and Kaplan, 1982) and their association to elevation, landcover and digital terrain information within the reporting of human perception poses an important challenge to the established manner in which the interrogation of landscape information is implemented.

Most studies attempting to assess landscape perception have employed questionnaire surveys which incorporate the comparison or compilation of significant landscape components (Potter and Wagar, 1971; Leopold, 1974; Countryside Commission, 1986). However, the manner in which the results are collated often make it difficult to relate the data to digitally defined information such as DEM values.

What we need to achieve

This paper seeks to investigate the cognitive and digital interface of landscape value assessment by examining several elements of landscape experience so that their inclusion in GIS may be facilitated. Before examining any associations between the physiographical and cognitive elements, it is suggested that a revised approach to the interpretation of the variables themselves needs to be adopted. This discussion is structured around *physiographical, specific feature, cognitive, and viewer interest variables* (Kliskey and Kearsley, 1993) and will generate suggestions that are specifically designed with GIS application in mind. Current GIS functionality could support the generation of some of these variables (Orland B,1992), for example, components that could be defined as *physiographical*, but it is not possible to derive a measure of landscape value from them. It is accepted that the choice of elements discussed is based upon the assumption that their inclusion will contribute to landscape cognition and provide the foundation for a series of indexed GIS variables. As part of their inclusion within GIS many of the implementations discussed assume some form of Boolean modelling which would dictate the *point* at which individual measures become significant or insignificant in landscape assessment. It is suggested that rather than seeking to ascertain specific points, an exploration of fuzzy membership functions needs to be undertaken which would be based upon context dependent values (Robinson,1988).

There are three principle aims:

a) to extract certain physiographical characteristics from the available digital data that are expected to contribute to landscape evaluation

b) to identify suitable cognitive criteria used in the subjective appreciation of landscape quality

c) to combine the extracted physiographical and cognitive elements in a GIS environment and subsequently, through functionality development, investigate their related roles in landscape value assessment.

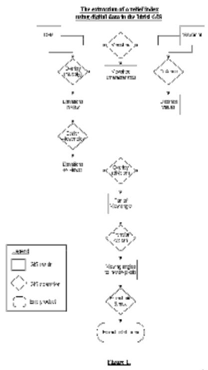
The structure of discussion in this paper will be in a question and answer format where challenges will be described with implementable and researchable answers suggested. It is accepted that in some cases, the indices created may prove less suitable in a GIS environment than others. The incorporation of cognitive elements and their interpretation still requires more research.

Physiographical characteristics

The landscape is the product of a multitude of related components which interact at a range of differing scales. For the purpose of this paper, *relief, depth-of-view, horizon characteristics and shape* are examined and GIS operations are suggested that could identify and analyse them as specific landform morphology components (Gobster and Chenoweth, 1989).

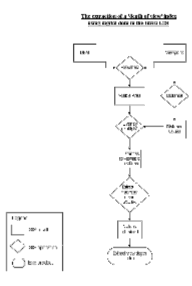
Q. How can we accurately assess the significance of *relief* in the viewshed?

A. Relief is an ambiguous concept that is generally considered to be a function of elevation (Mark, 1975; Evans, 1980). Using distance and the viewing elevation data in conjunction with relief angles, a measure of relief



may be derived which is sensitive to perspective (Fig.1). An answer may also be derived from the extraction of maximum and minimum viewing angles from a given viewpoint [eg. the line of sight operation *r.los* in the *GRASS* GIS]. The viewing angle may in itself be considered a simple measure of relief when examining a series of independent points, but may not be a sufficient indicator as individual elevation values do not allow for concepts such as the relative importance of landscape features. It is therefore suggested that a better indicator of relief is *volume.* It is considered that viewers of a landscape perceive quality of the view to be related to the amount of land or sky within the given viewshed. The volume of air-space within a particular view and the ratio of land to sky can be obtained from digital data and it may be possible to incorporate a weighted function to describe their importance within a given view. For instance, the inclusion of both two and three dimensional geometric characteristics of the view area could provide insight into the relief component by combining volume and perspective viewing within a digital analysis toolpack.

Q. How can we interpret the *depth of view* or the effects of perspective in a given viewshed?



A. It is simple to extract a summary *depth of* view value from the viewing angle function described above (Fig.2). However, the appropriate inclusion and significance of the incorporation of such a measure within landscape value assessment remains unclear as such analysis could produce several different components. For example, the distance of the furthest point from the viewer may be a high or low value depending on the shape of the viewshed and as a result viewer position has to be considered carefully in the interpretation of the landscape (Unwin,1975). An alternative approach may be to generate an area weighted mean value (from viewer to all points within the viewshed) or a standard deviation component for all such points. The application of a differential weighting of distances could prove effective in resolving problems surrounding the perspective component in such a model.

Q. In what way can the characteristics of the *skyline* and other intermediate *horizons* be incorporated within operational GIS functionality? Is it possible to quantify the way in which different types of horizons contribute to the view quality?

A. Although the skyline is obviously an important factor in the appreciation of the landscape, its extraction is not usually available within GIS functionality. If this was obtainable, its length and that of intermediate horizons could be associated in the analysis of the two-dimensional planimetric measures for a given viewshed area. An assessment of the contribution of linear horizon features could then be undertaken including, for instance, their density within the view and the area they screen compared to the viewshed area. Characteristics of each horizon such as their smoothness and the number of times the horizon is broken could also be incorporated which would provide the first steps to producing a measure of horizon dominance and the subsequent description of individual horizon qualities which may affect view quality.

Q. Can the ways in which different *view-shapes* affect cognitive assessment be parameterised? For example, are different views more pleasing than others as a result of their viewshape characteristics, and if so, can this association be measured when interrogating digital data?

A. Following on from the generation of a skyline and horizon function, a complementary component would enhance the characteristics of the local landscape as viewed from a point whilst reducing the importance of distant landscape components. This function would incorporate consideration of the linear boundary features and horizons of a viewshed and would be designed to emphasise the dominance of a given landscape element in the foreground whilst reducing the impact of an identical feature in the background (Craik,1972). A proportional scale could be applied to morphological features in the context of assessing their contribution to landscape value with local *outliers* and *extreme features* being an important consideration in this weighting function.

Presence of specific features

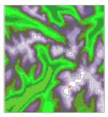
The established conventions applied to the analysis of specific landscape features combine generalised characterisations of a topographical surface with individual landscape elements (Hadrian *et al*, 1988). Whilst the shapes and forms of the world's surface can be modelled within the GIS environment it is not so simple to define the specific boundaries of *mountains* and *valleys*, *plains* and *plateaus* for digital analysis. In addition, questions remain as to the manner in which scale dependent classes of such relief components may be extrapolated. *Specific features* may be seen as *landsurface elements* such as mountains, valleys, plateaus, *natural landcover* for example rivers, woods, moors and *cultural features* including settlements, archaeological features and other human impacts.

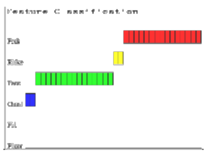
Q. How do we identify surface shape?

A. The process of systematically identifying surface form from mapped data has a long history dating from at least the mid-nineteenth century (Cayley, 1859). More recently, automated procedures for characterising surface form have concentrated on extraction from raster DEMs (eg, Evans, 1980; Pike, 1988; Meisels *et al*, 1995). One of the consistent problems observed with such extraction is that results are, in part, dependent on the scale and resolution implied by the raster data model (Evans, 1979; Hodgson, 1995; Polidori, 1995; Wood, 1996*a*). Any characterisation of surface form should be as independent of the data model as possible.

The approach adopted here to model surface form is to fit a bivariate quadratic surface through local 'kernels' passed over a regular gridded DEM. The derivatives of the modelled surface allow six morphometric feature types to be identified (Evans, 1979). These are, *pits* (local concavities in all directions); *channels* (local concavity in one direction, planar in an orthogonal direction); *passes* (local convexity and concavity in orthogonal directions); *ridges* (local convexity in one direction, planar in an orthogonal direction); and *peaks* (local convexities in all directions). Any part of the surfaces not identified in one of the categories above is regarded as *planar.* An algorithm for deriving such measures is detailed in Wood (1996*b*).

Q. How do we incorporate the uncertainty of feature definition?





A. Uncertainty in feature definition arises in part because the same location can be considered part of a number of different features simultaneously. While this may be for a variety of contextual reasons, it is asserted here that *scale* is a primary cause of this uncertainty. Landscape in the foreground of a view will inevitably be viewed at a contrasting scale to that which makes up a distant horizon. Equally an observer may consider a landscape as a whole, or be concerned with only small parts in greater detail. It is additionally possible that the very change of surface from with scale is itself an important part of a landscape view.

A procedural solution to this question is offered by deriving surface features over a variety of scales and recording not a single value, but a *feature membership function* for any one location. This is illustrated in Fig. 3 which shows the graphical output from an interactive interrogation of a DEM. Here, size of kernel increases along the X-axis by 100m increments. In this example, Mickledore which lies between Scafell and Scafell Pike in the English Lake District), is seen to be part of a channel at the finest scale (< 200m), a pass at intermediate scales (200m - 2km) and a peak feature at the coarsest scales (> 2km).

Q. How can the contribution of individual *artificial* features present in the view in question be identified and parameterised?

A. Although it is possible to link tabular information to features within a GIS, this functionality is not, in itself, able to assess specific feature relationships. Given that viewer focusing occurs, does the presence of such a feature preclude the importance of other landscape elements to the extent of reducing their contribution to the overall landscape? For example, given the visual dominance of a named Mountain Peak, does the viewer become sufficiently focused on *it* to the point of ignoring a pipeline running down its side, that in any other view would be considered an eyesore? By identifying specific features and naming them according to their physiographical characteristics, it is suggested that it should be possible to relate cognitive information to such features in an effort to assess the differences in perceived contributions of both micro and macro landscape components within the viewshed. Subsequent identification of the contribution made by the different components of *specific features* can then be analysed by simple database abstraction in addition to the perspective and scale criteria already defined by the viewshape characteristics.

It may also be possible to class the feature as a polygon with an assigned dominance value where the difference in the feature value to that of the surrounding landscape would provide a means of categorising it as integrated, intrusive, dominant etc.

Cognitive Criteria

The landscape architecture literature suggests that particular landscape components can be considered to have an effect on the quality of any view, e.g. water is generally considered a positive attractor (Leopold, 1969), whereas a road is considered to be negative (Potter and Wagar, 1971). In seeking to build upon such conclusions, a questionnaire has been devised whereby the responses obtained will be used to provide a means of assessing *how* beneficial or negative these affects may be. The structure of the questionnaire is designed to generate responses that will aid in the identification of certain cognitive elements that are considered particularly suitable to association with the digital data. These include *drama, mystery and coherence*. (Kaplan and Kaplan, 1982). In addition, trade off assessments may be evaluated through the utilisation of multi-criteria-decision-making models (Jankowski,1995) in an attempt to gauge relative impacts of specific features within the viewshed of interest (Higuchi, 1983). This type of feature description can then be related to the textural analysis within the digital environment and could provide a cognitively defined application to classify the landscape using harmony or chaos values. For example, is the accepted beauty of a scene compromised by the presence of a certain detractor to the extent that the view looses its beauty?

At present, none of these components can be addressed with GIS functionality, but it is suggested that the following questions may be proposed and subsequent answers explored through the combination of cognitive and digital data.

Q. It is anticipated that the presence of certain features make certain landscapes more *dramatic* to view than others. Is it possible to identify from the cognitive response data combinations of physiographical and geometrical components that may contribute to differing landscape quality and facilitate value analysis within a GIS?

A. It is expected that the viewers perception of the landscape can be related to the plan geometry characteristics derived from digital data. It is proposed that a viewing position from the top of a mountain is dramatic, but a view from the base of the same mountain peak could be similarly described. It may be possible to assess drama within a GIS by categorising the viewshed into proximal, intermediate and distant viewing areas and combining this element with the maximum and minimum viewing angle. For example:

i) In the proximal viewing region (0m - 1km from viewer) drama may be created by the presence of a cliff or precipice where the angle of relief is significantly greater than the viewing angle. This could be seen as particularly dramatic.

ii) In the middle region, (1km - 5km) drama tends to be created by the presence of a peak or significant visible topographic variation to the surrounding area. The viewing angle would be closer to the relief angle and the drama would then be derived from a combination of angle, feature and scale information.

iii) In the distant viewing area, (5km - skyline horizon) drama is created by a large-scale landscape feature such as a volcano or mountain range, and as a result, the impact of the viewing angle may be a lesser consideration. In this case, the skyline shape would be combined with viewangle and relief components.

Further investigation may identify geometrical elements that are of particular importance when relating the concept of drama to those of scale, perspective and more importantly the *relief* component(Litton, 1974). It is proposed that *drama* is a function of the corporate effects of physiographical, planimetric and cognitive criteria. The result of such investigation should provide the functionality to combine data from such categories to provide a cumulative descriptive index of landscape drama.

Q. It is suggested that the concept of *mystery* contributes greatly to landscape assessment but how can this contribution be assessed and then substantiated?

A. *Mystery* is closely linked to the distance that a viewer can see and the viewshape of the visible area (Gimblett *et al*, 1985). Typically, it is suggested that a view that has a *mysterious* component contains areas that the viewer knows must be present but are out of view and yet within relatively easy access. The concept of *discovery* is linked to that of mystery, and in the context of landscape assessment is usually made more acute by the presence of some form of access to the areas out of sight. For instance, in a valley scene, intrigue may heighten the concept of discovery by a pathway leading along its floor or, in a corridor view by the visibility of features beyond a local or intermediate horizon. In both examples, the viewer is drawn into the landscape by the intrigue of what lies ahead or within. By analysis of the horizon characteristics and masking of visible areas, it should be possible to generate a mystery component when combined with landsurface and landcover information. The generation of such an indicator will depend greatly on the ratio of land in and out of view within the viewshed and the areas lying beyond the horizons in question.

Q. *Coherence* is used to describe the fractal nature of the landcover within the visible area. How is coherence measured, and what effect does it have on the way in which people assess the quality of the landscape?

A. It is the contention of the authors that coherence affects how people feel and thus the manner in which cognitive responses are generated. For example, if a given area has a large number of landscape components that are highly spatially auto-correlated (negatively or positively) the viewer may be less likely to consider it attractive. This is because the landscape will appear visually uniform and possibly boring, whereas if the land surface elements are more randomly arranged, the viewer is more likely to identify areas of interest and variation (Crofts, 1975). Examples of the former can be found in regions dominated by conifer afforestation practices and large-scale grain crop planting. By composing a digital representation of the landsurface and landcover information, this suggestion can be ratified by combination with cognitive data obtained from respondents in areas of differing vegetation and cultural characteristics (LaGro, 1991).

How do we relate the cognitive and digital elements within a GIS environment?

Discussion has shown how absolute values may be extracted from digital information. It may then be possible to relate these values to cognitive components. To parameterise such relationships the questionnaire used provides the foundation for the interpretation of both physiographical and specific feature components within the data and specifically targets the appreciation of landscape components that can either be currently analysed or may form an element of newly proposed functionality. The cognitive data generated is expected to complement the physiographical characteristics discussed above. The respondents to the questionnaire are requested to interpret the viewed landscape according to its feature composition, attractors, detractors and potential user qualities. The subjectively defined criteria obtained are then combined with the digital data in an attempt to extrapolate the operational functions, as illustrated in Table 1.

Cognitive Data	**Digital Data**	**Operational Function**
the most satisfying place to view the landscape from	physiographic and relief assessment	satisfaction index
viewing the landscape with land above or below the viewpoint	line of sight and viewshed viewing angle	drama index
interpreting the similarities of the view to one of three photos	viewshape depending on combination of 2D & 3D data	visibility and pictorial control element
comparing the view to a series of digitally produced images	texture and pattern index of landsurface elements	coherence and aesthetic value
relating viewing position to a 2D illustration of the landscape	2D & 3D planimetric comparisons and relief element	perspective and scale measurement
measuring the impact of linear components within the view	linear features such as skyline and intermediate horizons	linear feature measurement
assessing view quality over considering entire viewshed	viewing angle and variable angles of relief within the view	view quality measurement
identification of visual attractors and detractors	specific feature dominance values and harmony component	specific feature contribution
inclusion intrigue and discovery as viewer is drawn into the view	view-shadow and ratio of horizons to visible area	mystery index

Table 1: The combination of cognitive and digital data to create GIS functionality

The relationships between cognitive and digital elements are examined by comparing the data obtained from the questionnaire and the digital information stored in landcover and DEM coverages. For example, it is believed that associations between the viewer position (Litton,1968) and the expected viewer satisfaction may be illustrated, and that the esthetic experience may be determined from a combination of the texture and pattern of the land cover information as and the digital plan form of the viewshed. When incorporating distance and scale variables, it is anticipated that if the viewer can accurately interpret the viewshed characteristics from a viewpoint, the sense of satisfaction and appreciation of the wider landscape will be enhanced. It is also the contention of this paper that there are relationships between the number and shape of the horizons present within a viewscape and the pleasure experienced by the viewer, as well as many other possible combinations of variables that may be interpreted.

A fourth category of landscape value assessment that has not been explored is the *interest* factor. This criteria is probably the most subjective of the landscape components and hence the hardest to generate absolute results for GIS inclusion. However, in the context of the above approach to landscape evaluation, it is possible to generate interaction measures from the questionnaire data by identifying feature classes of interest (Veal, 1974; Daniel and Boster, 1976). For example, a rock climber will be drawn to a view of a cliff from the base whereas a hang-glider will have more interest in the same feature from the top; an archaeologist's eye will be drawn to specific cultural features irrespective of other features when a hiker might be more interested in the location for its views. Measures of viewer interest can be addressed in the same way as specific features, with a weighting index depending on other elements such as access, facilities, and habitat.

This paper has illustrated some of the possible associations that may be explored between physiographical, feature and cognitive components in the discussion of the GIS environment. It remains to be seen which of the suggested components discussed above lead to enhance currently available GIS functionality. It is the conclusion of this paper that deterministic digital analysis in GIS can become more accurate by adopting an increasingly flexible approach to cognitive criteria in an effort to accommodate the subjective decision making processes that are currently overlooked within concept of Geographic Information systems.

References

Appleton, J. (1975) *The experience of landscape*. London: J Wiley and Sons

Bishop, I.D., Hulse, D.W., (1994) Prediction of scenic beauty using mapped data and geographic information systems. *Landscape and Urban planning* **30,** 59-70

Cayley, A. (1859). On contour and slope lines. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science.* **XVIII**, *264-268.*

Countryside Commission (1986) Wildlife & Countryside Act 1986, *Conservation maps of National Parks: CC no.**6**.*

Craik, K.H. (1972) Appraising the objectivity of landscape dimensions, *in* **Krutilla** (*ed*) Natural environments: studies in theoretical and applied analysis, 292-308, Resources for the future inc.

Crofts, RS. (1975) The landscape component approach to landscape evaluation *Transactions of Institute of British Geographers* **66,** 124-129.

Davidson D. A., Watson A. I., Selman P. H. (1993) An evaluation of GIS as an aid to the planning of proposed developments in rural areas. In **Mather, P.M.** ***(ed)*** *Geographical Information Handling: Research and Applications*, 251-259, London:Wiley

Daniel, T.C., Boster,R.S. (1976) Measuring Landscape Aesthetics: The scenic beauty estimation method. *USDA Forest Service Research paper* **RM-167**

De Floriani L, Marzano P, Puppo E. (1994) Line-of-sight communication on terrain models. *International Journal of Geographical Information Systems* **8**, 329-342

Evans, I.S. (1979) An integrated system of terrain analysis and slope mapping. *Final report on grant DA-ERO-591-73-G0040*, University of Durham, England.

Evans, I.S. (1980) An integrated system of terrain analysis and slope mapping. *Zeitschrift fur Geomorphologie*, Suppl-Bd 36,*.274- 295*.
Fisher P F. (1991) First experiments in viewshed uncertainty: The accuracy of the viewshed area. *Photogrammetric Engineering and Remote Sensing* **57**, 1321-1327
Fisher P F. (1993) Algorithm and implementation uncertainty in viewshed analysis. *International Journal of Geographical Information Systems* **7**, 331-374
Fisher P F. (1994) Probable and fuzzy models of the viewshed operation. In **Worboys, M.** (*ed*) *Innovations in GIS 1*, 161-175, London:Taylor & Francis
Fisher P.F. (1995) An exploration of probable viewsheds in landscape planning. *Environment and Planning B: Planning and Design* **22 (4),** 527-546.
Gimblett H.R., Itami R.M., Fitzgibbon J.E. (1985) Mystery in an Information processing model of Landscape preference. *Landscape Journal* **4 (2)**, 87-95
Gobster P H., Chenoweth R E. (1989) The dimensions of aesthetic preference: a quantitative analysis. *Journal of Environmental Management* **29**, 47-72
Hadrian D.R., Bishop I.D., and Mitcheltree R. (1988) Automated mapping of visual impacts in utility corridors, *Landscape and Urban Planning* **16**, 261-282
Higuchi T. (1983) *The visual and spatial structure of landscapes* [eg.p72] (translated by **Terry C.S.),** Tokoyo: Gihodo Pub.Co.Ltd.
Hodgson, M.E. (1995) What cell size does the computed slope / aspect angle represent ? *Photogrammetric Engineering and Remote Sensing*, **61(5)**, *513-517.*
Jankowski P. (1995) Integrating GIS and mulitiple criteria decision- making. *International journal of GIS* **9** (3), 251-273
Kaplan S., and Kaplan R. (1982). *Cognition and Environment: Functioning in an Uncertain World*, New York: Praeger
Kliskey A.D., and Kearsley G.W. (1993) Mapping multiple perceptions of wilderness in southern New Zealand. *Applied Geography* **13** 203-223
LaGro, J. (1991) Assessing patch shape in Landscape Mosaics. *Photogrammetric, Engineering and Remote Sensing* **57**(3) 285-293
Lange, E. (1994) Integration of computerized visual simulation and visual assessment in environmental planning. *Landscape and Urban planning* **30** 99-112
Lee J. (1991) Analyses of visibility sites on topographic surfaces. *International Journal of Geographical Information Systems* **5**, 413-429
Lee J. (1994) Visibility dominance and topographic features on digital elevation models. *Photogrammetric Engineering and Remote Sensing* **60**, 451-456.
Leopold L,B. (1969) Quantitative comparison of some aesthetic factors among rivers. *Geological survey Circ.* **630** 1-16 USDI G.S.Washington D.C.
Litton R.B.Jr. (1974) Visual vulnerability of forest landscapes *Journal of Forestry* **72** (7), 392-397
Mark, D.M. (1975) Geomorphometric parameters: a review and classification, *Geografiska Annaler* 57 A, 165-177.
Meisels, A., Raizman, S. and Karnieli, A. (1995) Skeletonizing a DEM into a drainage network, *Computers and Geosciences,* **21** *(1), 187-196.*
Miller D.R., Morrice,J.G., Whitworth,P.L., Aspinall,R.J. (1994) The use of GIS for the analysis of scenery in the Cairngorm Mountains *in* **Heywood and Price** *(eds*) *GIS in Mountainous Regions*, London: Taylor & Francis
Orland B. (1992) Data Visualization techniques in environmental management: a research, development and application plan. *Landscape and Urban Planning* **21**, 241-244.
Pike, R.J. (1988) The geometric signature: Quantifying landslide terrain types from digital

elevation models, *Mathematical Geology, 20 (5), 491- 511.*
Polidori, L. (1995) Fractal-based eevaluation of relief mapping techniques, *in* **Wilkinson, G, Kanellopoulos, I. and M�gier, J.** (*eds) Fractals in Geosciences and Remote Sensing*, Joint Research Centre, Report EUR 16092 EN.*277-297.*
Potter D.R., Wagar J.A. (1971) Techniques for inventorying manmade impact in roadway environments *USDA* FSRP **PNW-121** 12p
Preece R.A., (1991) *Designs upon the Landscape*. London: Belhaven Press
Robinson, V.B. (1988) Some implications of fuzzy set theory applied to geographic databases. *Comput., Environ. and Urban systems* **12**
Schafer E.L.Jnr, Hamilton J.F. Jnr, Scmidt E.A. (1969) Natural Landscape preferences: a predictive model. *Journal of Leisure Res.* **1**(1) 1-19
Unwin, K. (1975)The relationship of observer and landscape in landscape evaluation *Transactions of Institute of British Geographers* **66** 130- 134.
Veal, A.J. (1974) Environmental Perception and Recreation.A review and annotated bibliography
Wood, J.D. (1996*a*) The geomorphological characterisation of Digital Elevation Models, *PhD Thesis, University of Leicester, UK.*
Wood, J.D. (1996*b*) Scale-based characterisation of Digital Elevation Models, *in* **Parker, D.** ***(ed)*** *Innovations in GIS 3, Ch.14*, London: Taylor and Francis

---

All authors may be contacted at the address below,

Department of Geography
University of Leicester
Leicester LE1 7RH
UK

Telephone, fax and Email should be addressed to Dr. Peter Fisher on,

Phone: 0116 252 3839
Fax: 0116 252 3854
Email: pff1@le.ac.uk

David R Miller

Landscape Visualization using DEM data derived from Digital Photogrammetry

This paper discusses and illustrates the value of digital photogrammetric techniques combined with automatic characterization of landscape views for mapping the geographic distribution of view types. Digital photogrammetric techniques have been used to derive digital terrain data of areas as an alternative to digitized contour methods and the resultant surface analysed to produce a census of visibility.

INTRODUCTION

National initiatives on the production of digital datasets derived using digital photogrammetric techniques has been gaining ground for the provision of topographic basemaps, map updates or for interpretation of land ocver data (Goebel and Price, 1993; Gunnarsson, 1993; Light, 1993; Skalet *et al.* 1992). Currently, the most significant of these datasets is the ortho-photograph the value of which is for supplementing line maps, providing backdrop information to fill the `blank' areas on maps and therefore facilitating the user their own interpretation of features in the landscape. Such a source of data provides another input for use in Geographic Information Systems (GIS)(Chapman *et al,* 1991; Lenzen and Foresman, 1993).

The production of ortho-photography often utilizes elevation data obtained by conventional means, usually analogue or analytical photogrammetry, converted into digital form (Petrie, 1995). This approach has a higher level of productivity, avoiding as it does, the processing overheads associated with deriving the digital elevation model (DEM) directly from the same stereo models (Chapman *et al,* 1991Hood *et al,* 1989). In addition there are issues of the reliability of the elevation data which has been derived by such digital techniques (Petrie, 1995). The software for generating ortho-photographs on digital photogrammetric workstations, such as that produced by Intergraph, Zeiss, Autometrics and Helava is being complemented by that of Erdas Digital Ortho or Orthomax, HiView and PCI-Easipace and is becoming more reliable and housed on computing resources of greater power. Therefore, the generation of a digital terrain model directly from remotely sensed data is increasingly common (Konecny, 1979; Miller *et al,* 1992; Novak, 1992) and presents some of the same challenges of data quality as for contour or height point based products (Robinson, 1994), plus new ones, for example, feature matching (Graff and Usery, 1993; Li, 1993; Zheng, 1993) or deriving heights of feature (Howard, 1991).

One application area employing the use of both elevation data and physical and cultural topographic features is that of landscape planning, involving the perception of landscape and translating descriptors of the physical environment into models of the nature of the environment with which people may relate (Shafer and Brush, 1977; Daniel and Vining, 1983). This may include identifying the content of a view in terms of its vegetation and cultural features. The issue of the observer's ability to discriminate between features when viewed from a particular location means that techniques of distance queuing, hue attenuation and feature size and shape must be applied to that of perspective view geometry (McLaren and Kennie, 1989). The presence and distribution of surface features in the context of the topographic surface contributes to the type of view exposed to an observer (Mayall and Hall, 1994).

To place the calculation of view type into a wider geographic context, a census of visible land (Miller *et al*, 1995) is calculated, based upon a digital representation of the terrain. This provides a surface of land visibility with which land dominated by particular view types may be mapped (Miller *et al*, 1994). Land cover effects may be included in this analysis by adjusting the height of each location for example, approximate heights of build up areas or forests. The nature of this data is of a type which would appear to be provided for using digital photogrammetric techniques.

Study Area

The study area is in the south east of the Grampian Region in Scotland, from Glensaugh to the Cairn O'Mount (Figure 1). This area spans a significant geological feature, the Highland Boundary Fault. It is a low lying, relatively level agricultural land in the south and heather and peatland hill moorland in the north, ranging in altitude from approximately 100 metres to 456 metres above mean sea level.

Figure 1. Location of study area

The land cover within the area is approximately 9% agriculture, 17% forestry and 74% moorland (MLURI, 1993). Of the forestry, over 90% is plantation woodland, comprising mainly Sitka Spruce, Scots Pine and Douglas Fir. The 10% of broadleved trees is distributed around field boundaries plus a few individual trees growing in areas of rough grazings. The location is on a popular tourist route across the mountain, at the top of which there is a developed view point. Therefore, this area provides a range in the physical topography and land use against which to test the techniques for deriving elevation data and an associted census of visibility.

Derivation of Elevation Data

Three stereo-models were derived for the study area. The source photography was 1:24 000, panchromatic, near vertical with a focal length of 152 mm. AN AGFA flatbed scanner was employed in digitizing the photographs at a resolution of 1200 dpi. Topographic map information existed for this area at scales of 1: 10 000 and 1: 2 500 and an existing land survey control network (Miller *et al,* 1989), all of which could contributed control point data.

In addition, field surveyed control data was necessary in the moorland area to the north and west of the model area.

To provide the additional control a Trimble GPS was used in differential mode with a base station set-up at an Ordnance Survey triangulation pillar (tertiary control point in the national co-ordinated reference system of the United Kingdom). Fifteen control points were identified and their co-ordinates recorded. Nine of the additional points were identified in the upland, moorland areas of the study area, where no field control existed and no features are mapped on the base maps. The other six points were selected at features already mapped at 1:2 500 or 1:10 000 scales to clarify ambiguities of what the nature of features represented on the ground (such as the intersection of a fence or a ditch).

Scanner distortions were quantified using the measured co-ordinates of points on the scanned test grid compared to a computed grid intersections, from which a thrid order polynomial transformation was derived and applied to the original image (Burnside, 1979). This produced a scale distortion of 0.29 of a pixel at 1200dpi. The focal length, image coordinates and ground coordinates of ground control points were then used to determine the exterior orientation parameters of the image.

Control points, were selected for the model orientation. The sources were selected on the basis of clarity on the scanned photographs, relocation on the map or in the field and reliability of the control co-ordinates. Table 1 contains a summary of the source of control data and the accuracy with which it is recorded (measured or reported). Figures 2 and 3 show extracts of the the derived elevation model and the ortho-photograph for the area.

Control Source	Planimetric Accuracy	Height Accuracy
1:10 000	+-3 m	+-1 m
1:2 500	+-0.625 m	+-0.5 m
Ground Survey	+-0.1 m	+-0.05 to 0.2 m

Table 1. Control data (Height control data from the 1:10 000 scale map are Spot Heights, not contours.)

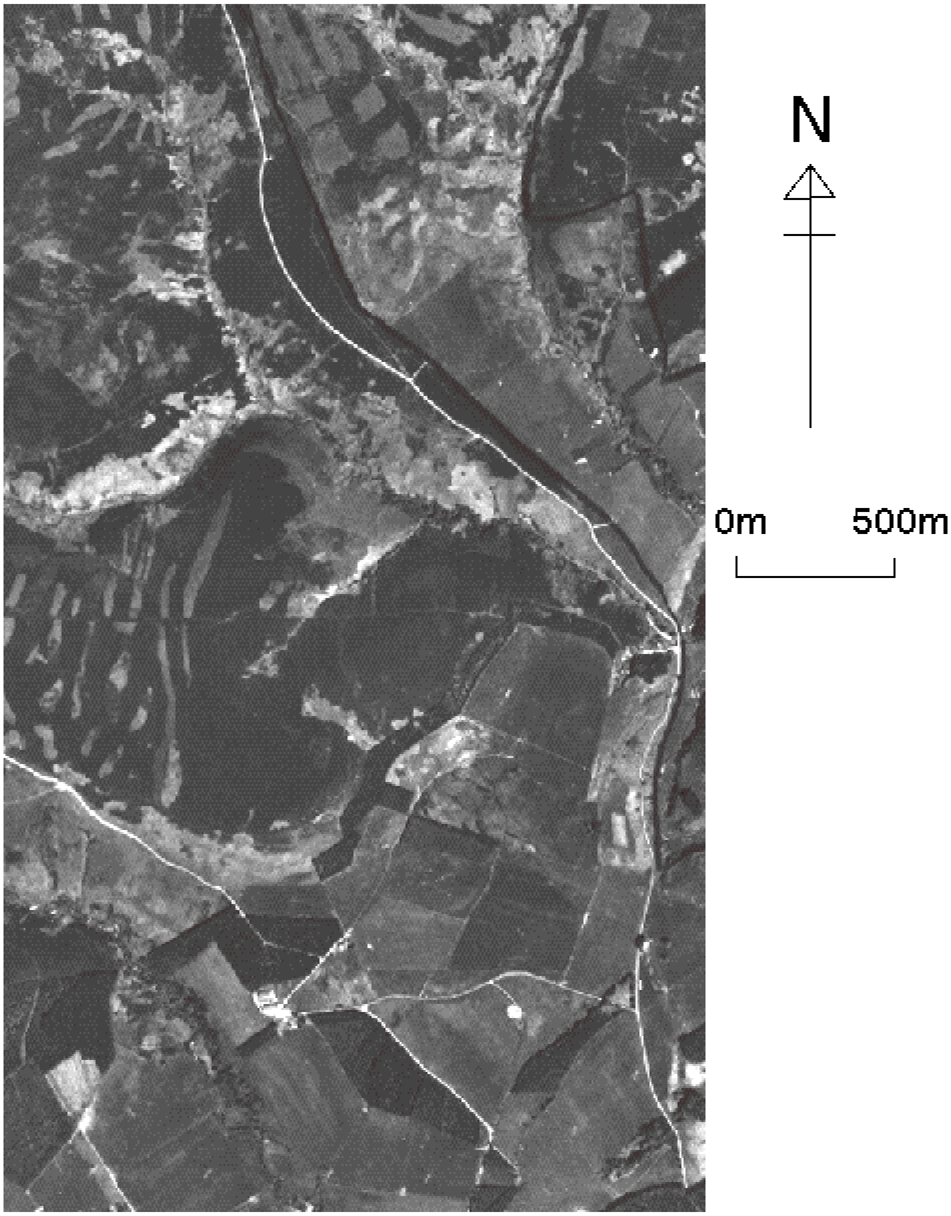


Figure 2. Ortho-photograph of an extract of study area.

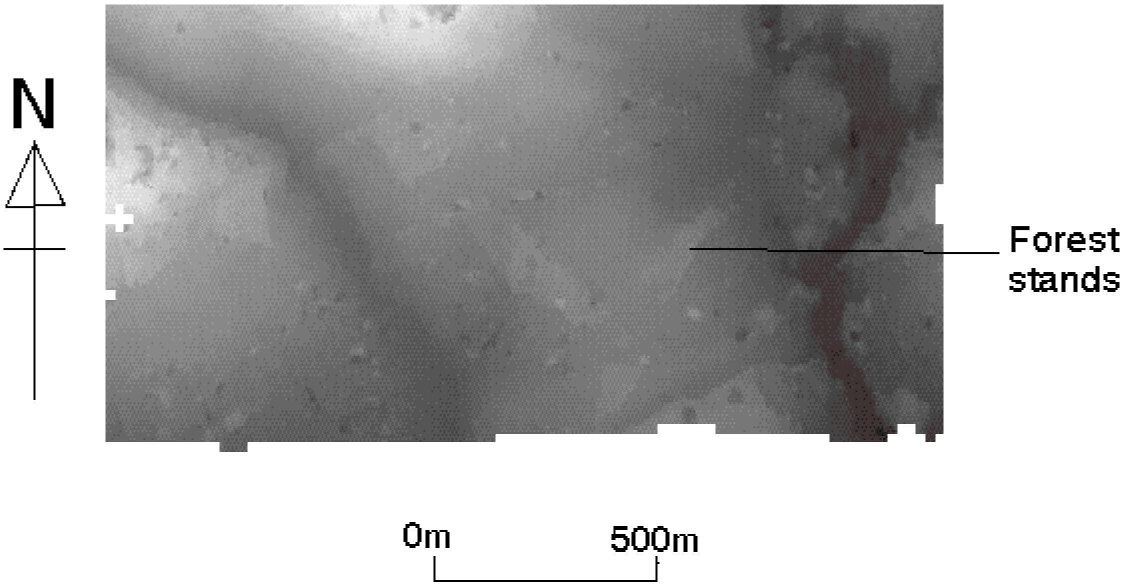


Figure 3. Digital elevation model of an extract of the study area

Check points were used to assess the planimetric and elevation accuracy of the derived products. The 11 check points comprised locations for which the co-ordinates were derived using traditional land survey techniques of theodolite and distance measuring equipment or global positioning systems (GPS). Table 2 contains a summary of the observations at the check points.

Number of Check points	Control Method	X-Coordinate RMS	Y-Coordinate RMS	Z-Coordinate RMS
6	Land Survey	+-0.13 m	+-0.11 m	+-0.15 m
5	GPS	+-0.36 m	+-0.45 m	+-0.83 m

Table 2. Geometric accuracy of model at check points.

Measurement of Tree Heights

Observations to trees within stands were restricted to those of which both the base and the top were visible. In certain plantations this necessitated the selection of sample points to be determined by the visibility of the trees. Thus, in plantations of narrowly spaced Sitka Spruce, which had not been thinned, there were fewer observations taken within the plantation compared to those taken for other species.

The height of individual trees was observed using two different optical or electro-optical devices: theodolite and electronic distance measuring equipment or inclinometer and tape measure (Philip, 1994). In each case the vertical angles to the base and top of the tree were observed and the slant distance to the base of the tree. From these observations, the height of each tree was derived, examples of which is included in Table 3.

Species	Field observation (m)	DEM observation (m)	Comments
Sitka Spruce	9.5	9.0	no thinning
Sitka Spruce	10.0	5.0	single line of trees
Sitka Spruce	8.35	8.0	no thinning
Sitka Spruce	6.2	6.0	no thinning
Ash	19.6	16.5	single tree
Ash	17.7	9.0	hedge
Scots Pine	14.0	15.0	close canopy
Scots Pine	8.15	8.0	no thinning
Scots Pine	8.8	8.0	thinned
Scots Pine	12.5	4.0	close canopy
Japanese Larch	19.1	18.0	row
Japanese Larch	13.2	12.0	thinned
Beech	20.0	19.0	single tree
Beech	17.5	16.0	single tree

Table 3. Example observations of tree height: field and DEM.

The observations were made six years after the original photography, thus a correction was required for their growth during the intervening period. The correction was for the growth during this period an indication of which was taken to be the number and spacing of the whorls at the top of the tree. Observations were made of the the height of the tree to the sixth whorl from the top and from this an estimate of the height of the tree at the time of the photograph was derived (Table 3).

The theodolite was used to calculate the difference in height between the bottom of the trees observed in the field and the observations to ground level of the open location. This ensured that the absolute difference in height between the point taken from the digitally derived elevation model and the tree and the top of the tree was known to an estimated accuracy of better than 0.25m. That is, to the order of magnitude with which one can ascertain what and where the base of the tree actually is.

The observations in Table 3 contain two examples of trees the height of which was not resolvable with reliability. One is an Ash and the second a Scot's Pine. In each case the tree was found in a narrow line of trees which were not adequately discriminated on the photography from the surrounding vegetation. As a consequence, the pattern matching algorithm was not able to remove the X-parallax in the local area and the derived heights were not reliable.

A linear regression between 80% (50 observations) of the field and terrain model observations on tree height give the following equation:

$$DEM\_Tree\_Height = 1.02 \times Field\_Measured\_Tree\_Height - 0.768$$

The RMS error for the regression is +-0.27 m.

The other 20% of observations have been used to validate the regression model. The standard deviation of the residuals from these additional points was 0.31 m.

The correlation between the measured and derived heights suggests that the methodology will be reliable for remotely measuring tree heights across large areas. The regression constant is likely to be attributable to the vertical resolution of the DEM being only 0.5 metre, the error in estimating the growth in the trees over the six year period and the observation being made at the edge of the stand or in gaps in the canopy.

Assessment of Visibility

The set of all locations visible from any specified location has been termed the *isovist* by Benedikt (1979) who described the importance of the observer's location as central because it is `representing the position of the observer whose spatial experience we are trying to explore'. The objective identification of the boundaries of this `spatial experience' is the basis for the following analysis of the digital terrain data. A census of the total area visible from all locations within the study area (each location is a pixel in the raster dataset) was calculated providing a relative intervisibility of land within the study area.

Figure 4 illustrates a perspective of the view up the mountain, draping the ortho-photograph across the derived DEM. The view is approximately north-west, with a range in altitude between the lowest (bottom right) and highest (top left) points of 300 metres. A road follows the shoulder of the hill across the centre of the illustration. The darker tones to the edges of the extract are predominantly heather moorland and the brighter tones on the valley sides are grasslands and bracken (*Pteridium aquilinum*).

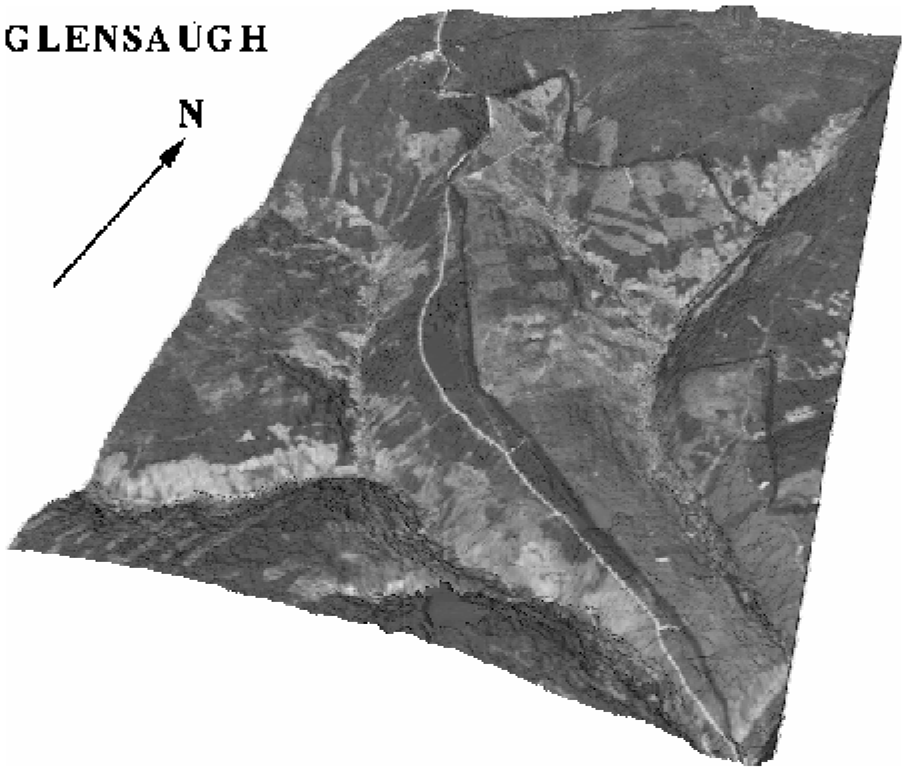


Figure 4. Perspective view draping ortho-photograph across the DEM

A calculation of intervisibility was undertaken for each cell and the total area visible was attributed to the cell. Each cell in the elevation model is counted only once, although the calculation could be weighted according to the inverse of the distance from the cell. The analysis considers a complete 360 degree rotation around each location for a radius of up to two kilometres on the raster DEM derived from the stereo-models. The output of this calculation is a cell-by-cell scoring of the visibility with a two km radius, based upon the topography derived from the three stereo-models, is restricted to a central area of 1.3 km x 5.5 km.

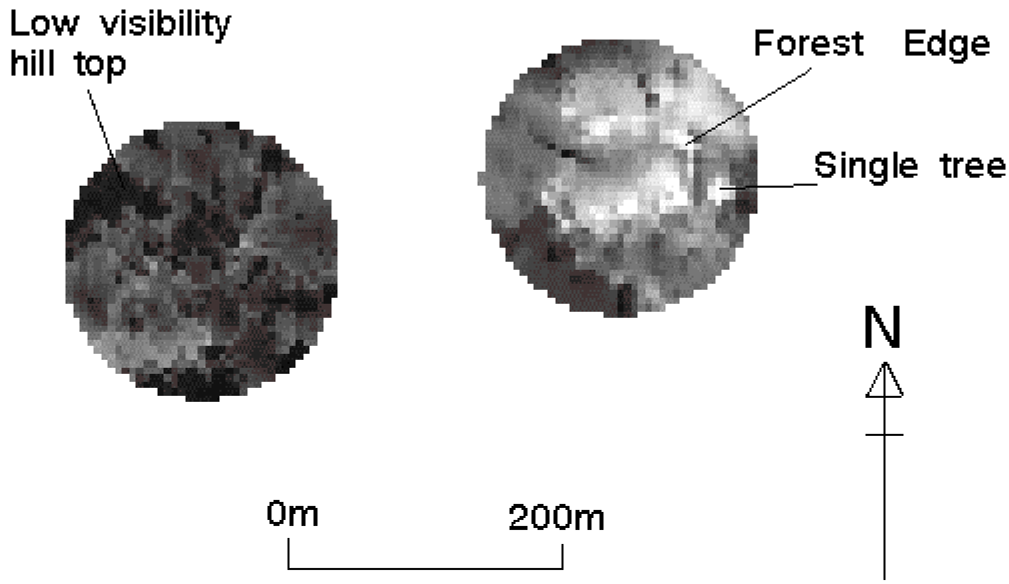


Figure 5. Extracts of visibility census.

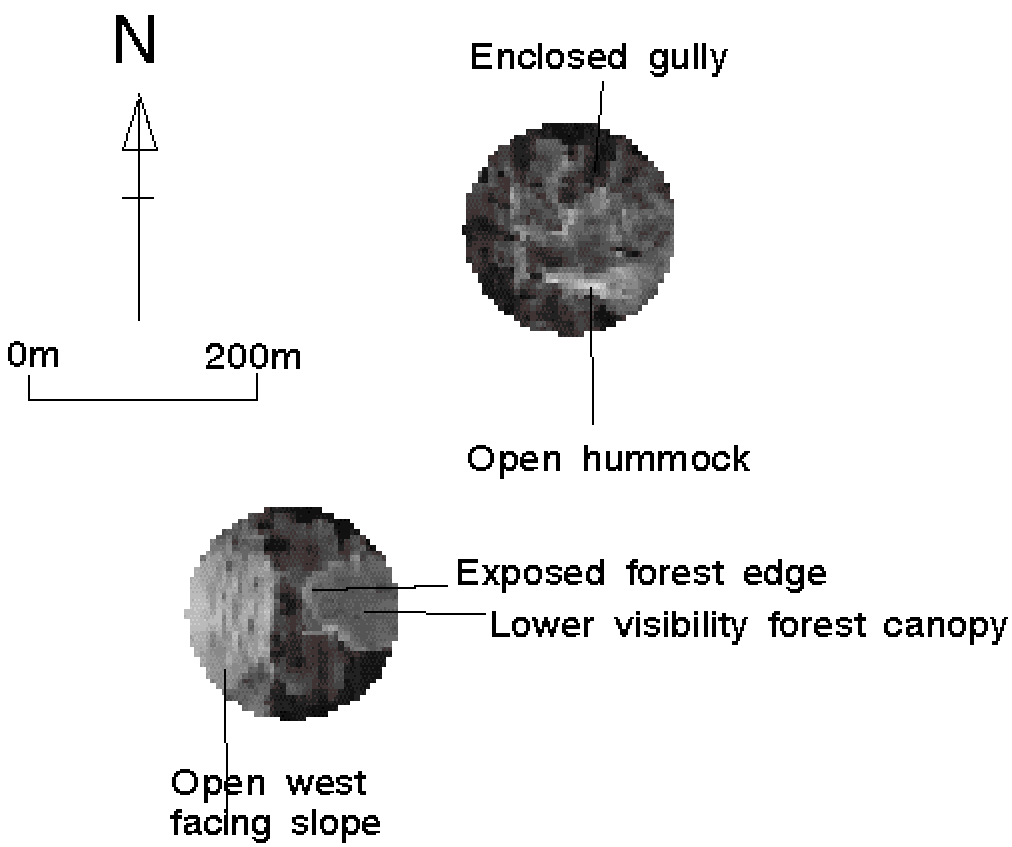


Figure 6. Extracts of visibility census.

Figures 5 and 6 are detailed views of two extracts of the area, showing the elevation data and the derived visibility census. Figure 6 has two areas highlighted, approximately 200 metres apart, with contrasting visibility levels and topographic features. On the left the land rises, with convex slopes, to a low hill top. The level hill top has a low visibility, with few cells having an open view of the land. On the right, there is a small stand of woodland, approximately 100 metres by 50 metres in size, the edges of which are highlighted as open to the surrounding land. The land to the south west has a low visibility level, on rising land which is hidden from the rest of the area by the forest stand.

Figure 6 shows two extracts from the south west of the area. At the bottom left there is an open, west facing slope which levels out and leaves the forest stand edge exposed compared to the surrounding terrain. The forest canopy is also measured as having as higher visibility than the surrounding terrain but lower than the stand edges. The example to the top right is open moorland where the variation in visibility levels is due to small, shallow gullies or low hummocks.

Overall, the visibility levels are influenced by:

1. extend to which the land is exposed to surrounding land; 2. the shape of the surrounding land;

The visibility of forestry is determined by the topographic context plus the height of the trees and their distribution across the terrain.

Combining the graphical summaries of the visibility in an area together with the observed values provides a means for identifying those areas of greatest prominance to an observer within an area and differences between the relative visibility of similar land cover features such as forestry.

Discussion

Despite over 25 years of the derivation of ortho-photographs they have played very little part in the increased use of GIS. Similarly, the increasing availability of digital photogrammetry and the potential for deriving digital elevation models has yet to be realized by a substantial percentage of the relevant parts of the GIS community (Brown and Bara, 1994). An increasing volume of data will become available across a continuum of scales, resolutions and precisional accuracies. Improvements in the quality of aerial photographic equipment and materials coupled with reduced costs of computer disk space and processing time will permit the expansion of a role of photogrammetry in GIS (Zilberstein, 1994) if the quality and flexibility of the products are matched to user need.

Further work is being undertaken on the measurement of within stand variation of tree heights using aerial photography from different years, and photography with and without trees present. Traditional methods of plotting contours using photogrammetric instruments necessitates a combination of interpolation between spot heights in areas of open canopy, and allowing a vertical offset for the heights of the trees and contouring the canopy top. Therefore, the contours in forested areas will necessarily be of lower accuracy than those in open terrain. Two opportunities may be taken to address this issue. Either the photography flown before planting for forest planning may be used or photography taken after felling so that if GIS were to be used in future planning an accurate DEM would be available. Thereafter, the monitoring of forest growth can be undertaken remotely.

The derivation of tree heights using digital photogrammetry is restricted by the same conditions that apply to aerial photographic interpretation and any photogrammetric measurements. For example, photographs flown in autumn, at a small scale, will not be of any value for deriving heights of broadleaved trees. Tree species will be further significant in that the shape of the crown will determine what is represented in the elevation model. Limitations will be:

1. Tree spacing too wide; 2. Tree height too small.

In each case, the trees (either individually or as a stand) have not been resolved in the derivation of the elevation model. Further analysis may be undertaken to present the uncertainty associated with the DEM based upon the check point observations. These observations, taken together with the patterns of residual parallax in the stereo-model may be used to derive a model of the uncertainty of the DEM with which one could modify the derived calulations of visibility (Fisher, 1994).

The calculation of terrain visibility has been artificially restricted by the use of a maximum scan of two km. In practice, the visibility of the terrain will extend significantly further. However, low resolution DEMs are inappropriate for the representation of surface features close to the observer, which will have a greater visual impact than those further away (McLaren and Kennie, 1989). Therefore, the resolution and detail should be tuned to match the distance from the observer, that is, the further from the observer the lower the resolution and the nearer the observer, the higher the resolution. The approach taken in this paper utilizes high resolution data, the analysis of which demands significantly higher computing resources than lower resolution data for the same area.

It would appear that further work is required on the representation of surface features for the analysis of visibility. Such work could utilize a data model which facilitates the accessing of records of features (such as trees or buildings) the visibility "shadow" of which could be computed with respect to the observer. Access to such records would be screened according to the likeliehood of being able to see such a feature at the location based upon the intervening topography.

The results from the analysis of visibility provides a method of comparing the visibility of locations on a comparable basis. If the analysis if undertaken over a sufficiently large area and with sufficient detail, the result is an absolute score for each location based upon the level of visibility of the land. Further refinement would require account to be taken of the nature of an object at the location. That is, the level of contrast of the object with its surroundings, the lighting conditions and its shape. An additional aspect would be any associated information that the feature and the prevailing conditions provided the observer which may enhance or reduce the level of significance of its presence, such as the casting of shadows.

In conclusion, the example presented indicates the extent to which digital photogrammetric techniques can provide additional data for use in landscape management and planning. The use of aerial photography as a source of the data permits updating of the estimates of height of the forest stands and thus monitoring of the changes in visual impact of the forestry within an area over time. Considerable additional work requires to be undertaken to translate the techniques and processing into an operational facility but complementary work in the fields of design and psychology will contribute to a more robust means of translating level of visibility into an appraisal of visual impact.

ACKNOWLEDGEMENTS

The author would like to acknowledge the assistance of Gary Wright and Richard Aspinall for discussion on the content of the work, Paul Finch and Charles Fon for their work on the study site and Alistair Law for his assistance with the computing requirements of this project. Funding of the project was from the Scottish Office Agriculture, Environment and Fisheries Department.

REFERENCES

Benedikt, M. L. (1979) To take hold of space: isovists and isovist fields, *Environment and Planning B*, 6, pp 47-65.

Brown, D G and Bara, T J (1994), Recognition and reduction of systematic error in elevation and derivative surfaces from 7.5 minute DEMs. *Photogrammetric Engineering and Remote Sensing*, Vol. 60(2), pp. 189 - 194.

Burnside, C D (1979), *Mapping from aerial photographs.* Fletcher and Sons, Norwich. pp. 304.

Chapman, D, Dowman, I and Muller, J-P (1991), Digital photogrammetry - interfaces with GIS. In: *Proceedings of the AGI'91 Conference*, Birmingham, November 1991.

Daniel, T.C. and Vining, J. (1983) Methodological Issues in the Assessment of Landscape Quality. In: *Behaviour and the Natural Environment* (eds. Altman, I. and Wohwill, J.),Chapter 2, 39-83, Plenum Press.

ERDAS Inc. (1992), *Digital Ortho Manual*, ERDAS Inc, Atlanta.

Fisher, P.F (1994), Probable and fuzzy models of the viewshed operation, In: *Innovations in GIS 1* (ed. M.F. Worboys), Taylor and Francis, pp 161 - 176.

Goebel, J J and Price, L C (1993), National resources inventory and digital ortho-photography databases: characteristics, availability and use. In: *The Proceedings of the 2nd International Conference on Environmental Modeling and Geographic Information Systems*, Breckenridge, Colorado, September 1993.

Graff, L H and Usery, E L (1993), Automated classification of generic terrain features in Digital Elevation Models. *Photogrammetric Engineering and Remote Sensing*, Vol. 59(9), pp. 1409 - 1417.

Gunnarsson, R J (1993), Digital photogrammetry: the foundation for Iceland's geographical database. *GIS Europe,* November 1993, pp. 37 - 39.

Hood, J, Ladner, L and Champion, P (1989). Image processing techniques for digital ortho production. *Photogrammetric Engineering and Remote Sensing,* Vol.55(9), pp. 1323 - 1329.

Howard, J A (1991), *Remote Sensing of Forest Resources.* Chapman and Hall, London. pp.

Konecny, G (1979), Methods and possibilities for digital differential rectification. *Photogrammetric Engineering and Remote Sensing,* 45 (6), pp. 727 - 734.

Lenzen, T W and Foresman, T W (1993), Digital image databases support GIS operations. *GIS World,* November 1993, 36 - 38.

Li, Z (1993), Mathematical models of the accuracy of digital terrain model surfaces linearly constructed from least square gridded data. *Photogrammetric Record,* 14 (82), pp. 661 - 674.

Light, D L (1993), The national aerial photography program as a Geographical Information System resource. *Photogrammetric Engineering and Remote Sensing,* 59 (1), pp. 61 - 65.

Mayall, K. and Hall, G.B. (1994) Information Systems and 3-D Modeling in Landscape Visualization. In: *Urban and Regional Information Systems Association Annual Conference Proceedings,* Vol 1, 796-804.

McLaren, R. A. and Kennie, T. J. M. (1989) Visualisation of digital terrain models: techniques and applications. In: *Three dimensional applications in geographic information systems,* (ed. J. Raper), pp. 79-98. Taylor and Francis: London.

Miller, A B, Helava, U V and Helava, K D (1992), Softcopy photogrammetric workstations, *Photogrammetric Engineering and Remote Sensing,* 58 (1), pp. 77 - 83.

Miller, D R, Morrice, J G and Whitworth, P L (1989), The bracken problem in Scotland. In: *Bracken'89: biology, control and management.* (eds J A Thomson and R T Smith), Australian Institute of Agricultural Science Occasional Publication No. 40. pp 121 - 132.

Miller, D R, Morrice, J G, Whitworth, P L and Aspinall, R J (1994), The use of GIS for the analysis of scenery in the Cairngorm Mountains, *GIS in Mountainous Regions,* (eds. I Heywood and M Price), Taylor and Francis.

Miller, D R, Law, A N R and Brooker N A (1995), Calculation of a census of visibility. In: *Procceedings of the 15th ESRI User Conference.* ESRI Inc, Redlands, Ca. USA. 1995.

MLURI, 1993, *The Land Cover of Scotland by aerial photographic interpretation.* A report to the Scottish Office Environment Department. The Macaulay Land Use Research Institute, Aberdeen.

Novak, K (1992), Rectification of digital imagery, *Photogrammetric Engineering and Remote Sensing.* 58 (3), pp. 339 - 344.

Petrie, G P (1995), Air photo inputs to GIS for environmentl and strategic planning - technologies, procedures and products. In: *Proceedings of Aerial photographs for strategic planning*, Remote Sensing Society GIS Special Interest Group meeting, November 1995.

Philip, M S (1994), *Measuring trees and forests.* 2nd Edition, CAB International.

Robinson, G J (1994), The accuracy of digital elevation models derived from contour data. *Photogrammetric Record*, Vol.(14)83, pp. 805 - 814.

Skalet, C D, Lee, G Y G and Ladner, L J (1992), Implementation of softcopy photogrammetric workstations at the U.S. Geological Survey. *Photogrammetric Engineering and Remote Sensing*, 58(1), pp. 57 - 63.

Shafer, E.L. and Brush, R.O. (1977) How to measure preferences for photographs of natural landscapes. *Landscape Planning*, 4, 237-256.

Tilley, G B (1994), Digital ortho-photography for natural resource management, In: *The Proceedings of the GIS'94 Symposium*, Vancouver, February 1994.

Zheng, Yong-Jian (1993), Digital photogrammetric inversion: theory and application to surface reconstruction. *Photogrammetric Engineering and Remote Sensing*, Vol LIX (4), pp. 489 - 498.

Zilberstein, O (1994), National GIS from aerial photographs using a softcopy photogrammetric workstation. In: *The Proceedings of the 4th European Conference on geographical Information Systems*, Paris, March 1994.

David R Miller
Macaulay Land Use Research Institute
Craigiebuckler
Aberdeen
Scotland
AB9 2QJ
Tel: +44 1224 318611 x 2240
Fax: +44 1224 311556
Email: d.miller@mluri.sari.ac.uk

Peter Berger, Paul Meysembourg, Jim Sales, and Carol Johnston

TOWARDS A VIRTUAL REALITY INTERFACE FOR LANDSCAPE VISUALIZATION

ABSTRACT

A prototype virtual reality interface which combines the visualization capabilities of virtual reality with the spatial display capabilities of GIS is described. Satellite-derived landcover databases are draped over a DEM-derived topographic frame for EPA's Mid-Atlantic Integrated Assessment (MAIA) area. Trees are depicted as three-dimensional objects, corresponding with forested covertypes, and other land covers are represented by different colors and textures. The user can specify a vehicle (e.g. airplane, all-terrain machine) and a path for travel, viewing landscape features from different perspectives.

INTRODUCTION

Geographic Information Systems depict spatially-distributed data as they would be shown on a map: two-dimensional surfaces viewed from nadir via a high platform, with spatial objects represented by different patterns and colors. However, humans located in an actual landscape view features much differently: the land surface is undulating, vegetation is three-dimensional and has characteristic structures, objects appear smaller in the distance, and features are located above, below, and around the observer. Consequently, many people find it difficult to visualize the data represented by a GIS.

We are developing a software interface which combines the visualization capabilities of virtual reality with the spatial display capabilities of GIS. Landscapes are rendered as perspective views using actual elevation and land cover data, such that they depict realistic scenery. The user can specify a vehicle (e.g. airplane, all-terrain machine) and a path for travel, viewing landscape features from different perspectives. We use the term "virtual reality" to denote a system which provides the tools for users to interact with a simulated environment, but not necessarily in real time.

The ultimate goal of the research is to develop a prototype virtual reality interface with data from the Mid-Atlantic Integrated Assessment (MAIA) region, in support of the U.S. EPA's Environmental Monitoring and Assessment Program EMAP). The MAIA region falls in the states of PA, VA, WV, DE, NJ, and MD, and the District of Columbia, and includes the entire Chesapeake Bay watershed. Virtual rendering is being accomplished with two basic datasets: (1) digital elevation models (DEMs), and (2) land cover data derived from classified satellite imagery. Specific technical objectives are:

1. Interface DEM with virtual reality software.
2. Drape classified land cover over DEM.
3. Set camera location, elevation, and view angle.
4. Render image, enabling 3-D structures.

METHODS

Software

In order to provide results quickly, we used an existing, inexpensive (~$100 US) virtual reality program, Vistapro 3.13 for DOS (Virtual Reality Laboratories, Inc., 2341 Ganador Ct., San Luis Obispo, CA 93401), which we ran on a Pentium P5-133 Intel based PC running DOS 6.2, with 16 mbytes of ram and 2 gigabytes of hard disk storage. Versions of Vistapro are also available for PC Windows and Macintosh platforms. Vistapro is a software package for three- dimensional landscape simulation that can use U.S. Geological Survey Digital Elevation Model (DEM) files to accurately recreate real world landscapes. It is a single frame generator, meaning that it acts like a camera; at intervals along a user-specified flight path, Vistapro points and clicks the camera, rendering a new view of the landscape. The individual frames are displayed in rapid succession to make "fly-bys" appear animated. The Vistapro package also comes with a freely distributable "player" program which allows a virtual world developer to make copies available to users who do not own Vistapro. The player provides capabilities for exploring the virtual world but not modifying it.

Vistapro renders realistic scenary by level slicing the landscape color palette to realistically shade cliffs and hills. It does this through user definable variables, a rules system, fractals, and chaotic math to add non-uniform textures to the rendering (Virtual Reality Laboratories, 1993). Lifelife landscapes are depicted despite a limited number of possible terrain types: water, beach, vegetation, bare, snow, cliff, and buildings.

Input data for defining landscapes can be DEMs and/or PCX files. PCX is a standard graphic file format commonly used by Paint programs in Windows applications. Landscapes can be built using only DEMs, only PCX files, or a combination of both. When only DEMs are used, terrain types are defined based on elevation (Fig. 1). For example, low, intermediate, and high elevations could be defined as water, vegetation, and snowfields, respectively. Alternatively, the white and tan colors of snowfields and bedrock on a PCX- format aerial photo of a mountaintop could be used to define those areas as user-defined elevation ranges, as well as terrain types. When both DEM and PCX files are used, the DEM is used to generate topography, and the PCX file is used to generate land cover.

Although Vistapro generally worked well for our purposes, importing GIS data into Vistapro was not straightforward, requiring the development of special protocols and the use of other software packages to pre-process the data (see Results).

Datasets

Initial testing and rendering was done with a DEM supplied by Vistapro of Crater Lake, Oregon, which allowed us to quickly generate realistic pictures and fly-bys of the region. After perfecting procedures for importing and exporting the supplied DEM, we learned how to import and export eight 7.5 minute USGS DEM's of Voyageurs National Park, which were not native Vistapro files and therefore posed additional challenges. Use of these data allowed us to develop procedures for joining adjacent DEMs to render larger and larger regions, a capability required to display scenes from throughout the multi- state MAIA region. Ultimately, we used four 1:250,000 USGS DEMs centered over the MAIA region

(Williamsport East, Harrisburg East, Scranton West, Newark West), which were obtained via ftp from the U.S. Geological Survey Home Page (http://www.usgs.gov/). Each 1:250,000 file covers a 1 x 1 degree block at a data resolution of approximately 90 x 90 meters at this latitude.

Two types of land cover data, both derived from satellite imagery, were draped over the DEM-generated landscape: (1) the Conterminous U.S. AVHRR Global Data Set, and (2) the Chesapeake Drainage Basin Land Cover Grid. The first data set was produced by the U.S. Geological Survey, and consists of 1-km resolution data derived from AVHRR satellite imagery, classified into vegetation types based upon spectral reflectance values and seasonal onset of greenness (Eidenshink 1992). The second data set was produced at the EPA EMSL-Las Vegas Laboratory from Landsat TM data (30 x 30 m pixels) of the Chesapeake Bay region, and consisted of six land cover classes: high-density developed, low- density developed, woody, herbaceous, exposed land, and water.

RESULTS

Vistapro's strength was that it could render beautiful images quickly, but its weaknesses were that: (1) adjacent DEMs joined with the MCNV program in Vistapro were slightly misaligned, and (2) Vistapro doesn't use true geo-referencing, which would prevent combining land use patterns with the DEMs (Objective 2). Therefore, we developed procedures using auxiliary software programs to overcome these deficiencies (Table 1, Fig. 2).

Table 1. Software packages used.

Software	Ver.	Platform	Function
ARC/INFO	7.0.3	Sun Sparcstation	Merge multiple DEMs
ARC/INFO Grid	7.0.3	Sun Sparcstation	Georeference, resample, clip
ERDAS Imagine	8.2	Sun Sparcstation	(same functions as ARC/INFO GRID)
XV	2.21	Sun Sparcstation	Manipulate file colors
Graphics Workshop	1.1	PC Windows 3.1	Convert TIFF to PCX format
SAGE Capture	2.13	PC DOS	Convert ARC/INFO DEM to Vistapro format
Makepath	3.10	PC DOS	Generate paths & camera angle
Vistapro	3.13	PC DOS	Render landscapes, generate fly-bys
Excel	5.0	PC Windows 3.1	Edit scripts

Objective 1: Interface DEM with VR Software

The Vistapro MCNV program aligns adjacent files using a row and column match, which caused seams to be visible in joined DEMs. Because of this misalignment problem, we used the LATTICEMERGE procedure in ARC/INFO (Environmental Systems Research Institute, Inc., Redlands, CA) to join adjacent DEMs. Although this worked well for splicing and aligning the DEMs, there were problems importing ARC/INFO-generated DEMs into Vistapro, because ARC/INFO's DEM output differs from the format used by USGS. Vistapro will read an ARC/INFO DEM, but the "nodata" values trick the program into interpolating between -9999 and the highest value in the file. This distorts features such as sea level and snowline, and corrupts Vistapro's use of level slicing. Level slicing is used in the rendering phase to give computer images a photorealistic quality. Therefore, a third software program, "SAGE Capture" (Digital Land Systems Research, P.O. Box 4191, Parkville, Victoria 3052 Australia) was used to convert the ARC/INFO-spliced DEMs into the correct format. The SAGE Capture program reads the ASCII file exported by ARC/INFO and can export either a Vistapro binary file or a Vistapro version of a DEM.

Objective 2: Drape classified land cover over DEM

Although Vistapro renders lifelike landscapes using elevation alone by assigning different vegetation types to user-specified elevation ranges (Fig. 1), we sought to vegetate the MAIA landscapes based on actual land cover distribution, which is often independent of elevation. We are developing procedures for interfacing land cover and DEM data in Vistapro, work that is still in progress.

Files in PCX format with up to 8 colors can be used to define terrain types in Vistapro, each color representing a different terrain class. Vistapro uses these classes to generate shading patterns and 3-D objects associated with particular vegetation types. To ensure that the colors in the land cover files corresponded with the desired terrain type, we converted them to a TIFF format and imported them into XV, a UNIX graphics utility, for color manipulation. After color editing, we ported the files to a PC and converted them from TIFF to PCX using Graphics Workshop, a shareware graphics conversion utility that runs under Windows 3.1.

Files to be overlaid had to be aligned before importing them into Vistapro. Vistapro uses row and column addresses rather than true geo-referencing, so it was necessary to create DEMs and land cover files covering exactly the same area and with exactly the same number of rows and columns. There were several steps involved in this procedure.

The first step was to georeference the DEMs and land cover datasets. We used ARC/INFO GRID to perform this georeferencing, but also had excellent results with ERDAS IMAGINE (ERDAS, Inc., 2801 Buford Highway, NE, Suite 300, Atlanta, Georgia 30329-2137), particularly for scanned aerial photographs or USGS 7.5' topographic maps.

The second step was to resample the files so that the cells in each dataset were of the same size and aligned exactly. The cells in the 1:250,000 DEMs were approximately 90 x 90 m, whereas the cells in the Landsat TM-derived land cover layer were 30 x 30 m. This was also done using ARC/INFO GRID.

The third step was to change the nodata values (-9999) of the DEM file to a value equal to one less than the lowest elevation in the DEM. If this is not done, Vistapro uses the -9999 value as elevation, thus severely distorting the landscape.

The fourth step was to clip both datasets to one of four standard sizes used for display by Vistapro. This step was essential because Vistapro displays DEMs and PCX files differently: DEMs are centered within the display screen, whereas PCX files are justified with the lower left corner of the display window. Vistapro pads the edges of files with 0 values if the rows and columns don't match one of its four standard sizes, which also causes misalignment of overlaid datasets. The clipping was done with the row and column references used by Vistapro, rather than actual locational information. After these procedures were done in ARC/INFO GRID, we exported the DEM into an ASCII format, and passed it through the import/export process in SAGE Capture (see Objective 1).

Objective 3: Set camera location, elevation, and view angle

The Makepath Flight Director (Virtual Reality Laboratories, Inc., 2341 Ganador Ct., San Luis Obispo, CA 93401) was used to generate the paths used to "fly" though the rendered landscape. Paths can be generated interactively or by use of scripting controls, which allow creation of multiple unattended views of a landscape. A DEM was loaded, and a flight path set by placing nodes along the desired route.

The script file generated by Makepath is a comma delimited ASCII file with one record per frame defining camera x, y, and z coordinates, heading, pitch, and angle of view. Vistapro uses the scripts to generate individual frames which can then be played back as an animation. Makepath permits only a limited range of motion consistent with the vehicle type chosen, and uses the elevation of the DEM to determine the elevation of the path for every frame.

The ASCII script file that Makepath generates can be edited to vary elevation along the chosen route or to simulate other types of motion. We used the Microsoft Excel 5.0 spreadsheet program to edit the script file, simulating a parachute jump into Crater Lake and a space capsule zooming into an entire 1:250,000 DEM quad. Neither of these vertical drops are inherent to Makepath. Several iterations were usually needed to fine tune the desired path. At this trial and error stage, low detail and course resolution were used to get a feel for the visual results of the fly-by because of the long amount of time required for final rendering (see Objective 4). Script editing can also be used to reduce rendering time by decreasing the detail and quality of files produced.

Objective 4: Render image, enabling 3-D structures

Scenes can be rendered in Vistapro at different levels of resolution. High resolution images are more aesthetic but require longer rendering times than low resolution images (Fig. 3).

Computer performance was an important determinant of rendering speed. Although the minimal hardware requirements for the DOS version are a 386 with a VGA display, 4 Megabytes of RAM and 3 Megabytes of hard disk space, more powerful equipment provided much faster results. Rendering speed benchmark tests were performed using the Vistapro supplied DEM of Crater Lake (258 x 258 cells) on three different computer configurations (Table 2). Display modes were all set at 8 bit VGA 320 X 200 resolution. Results of the rendering speed benchmark tests are shown in Table 3.

Table 2. Computers used in benchmark tests. Speed tests were done with the Norton Integrator Software Advanced Edition, Version 4.50.

Machine	System Info. Index	Disk Speed	Disk Transfer Rate
386 20MHz 4MegRAM	13.2	22.4ms	494 K/S

486 66MHz 8Meg RAM	144	11.5ms	1.4MB/S
P5 133MHz 16Meg RAM	420	8ms	3.6MB/S

Table 3. Results of rendering speed benchmark tests. Times shown are in the following machine order: 386/486/P5. All times are rounded to the nearest second.

	Texture Settings			
# of Polygons	**Off**	**Low**	**Medium**	**High**
2048	0:03/0:01/0:00	0:12/0:01/0:00	0:39/0:06/0:01	1:45/0:18/0:05
8192	0:08/0:01/0:00	0:22/0:02/0:00	0:45/0:07/0:01	1:52/0:20/0:05
32,768	0:24/0:03/0:00	1:02/0:09/0:02	1:17/0:12/0:03	2:13/0:23/0:06
131,072	1:48/0:14/0:03	2:50/0:22/0:06	2:53/0:25/0:06	3:21/0:30/0:08

Vistapro is more computationally intensive with large files than with the small file used in the benchmark tests. Rendering of large files (2050 x 2050 cells) ranged from three seconds to 13.4 minutes per frame on the Pentium P5 computer. A number of variables manipulated within the program itself also influenced rendering speed and quality: size of polygons, texture quality, tree visibility, blending, dithering, pixel dithering, and gouraud shading. Final processing of all frames took as long as eight hours for a 200 frame animation lasting 19 seconds.

The ability to render very accurate 3-D structures was important in our decision to use this software package. Fractal generated trees were used with various controllable settings such as: light source and direction, density, size, detail, and colors. These also affected the time required to produce the final output. Generating many fractal trees was extremely time consuming, but the visual results were well worth it (Fig. 3).

Enhancements

In addition to making progress on our basic objectives, we also developed the ability to edit individual frames of an animation to add interesting visual enhancements. For a test, we used a scanned image of a pair of boots to add realism to a simulated parachute jump over the Crater Lake Area. The frames were individually edited using a basic paint program. These individual frames were then re-processed into animation form for viewing. The result appears remarkably real, as if the parachute jumper is looking down at his/her feet! This process more appropriately could be used to add an inset locational map or data tables about the scenes currently being viewed during a fly-by. It should be noted that only 12 frames were edited for this test; editing all or many frames from a 200 frame animation could be very tedious, so an automatic file editor would need to be written.

CONCLUSIONS

Despite its fairly limited feature set, Vistapro allowed us to get a working prototype up and running quickly. We used the prototype both to visualize the land cover data and to benchmark performance requirements and discuss enhancements for future implementation.

Virtual reality renderings are computationally intensive, so it was not possible to make our applications interactive in real time with devices such as a headset or data glove. As the power available on consumer-grade desktop computers increases, however, progressively more realistic presentations become possible using ubiquitous hardware. These hardware advances have enabled a generation of low cost visualization and virtual reality software packages for the experimenter. Although the results of these low cost solutions only approximate the realism (measured in rendering detail and speed) available on high-end platforms, some are now good enough to enable serious work, and the results obtained can be improved markedly simply by moving the software to faster hardware as it becomes available.

At the extreme, virtual reality promises a fully immersive environment which fully engages all the user's senses and provides full interactivity in real time. At the minimum it provides a visualization environment that encourages a user to imagine being within the virtual world -- to suspend disbelief in the simulation. At this stage in our project, user interaction is limited to importing digital elevation and land use data and navigating within the resulting environment. Once we complete experiments with our prototype, we plan in later stages to port our model from Vistapro to a more open-ended virtual reality software package, which will allow us to effect changes to the model from within the virtual environment.

ACKNOWLEDGEMENTS

Support from Minnesota Technology, Inc., the U.S. Environmental Protection Agency EMAP Program, and the U.S. Department of the Interior, National Biological Service is gratefully acknowledged. No formal endorsement of the commercial software products mentioned should be inferred.

REFERENCES

Eidenshink, J.C. 1992. The 1990 conterminous U.S. AVHRR data set. Photogrammetric Engineering and Remote Sensing 58:809-813.

Virtual Reality Laboratories, Inc. 1993. Vistapro User Guide, IBM DOS Version.

AUTHOR INFORMATION

Peter Berger
Proprietor
Brimson Laboratories
1549 Hiironen Road
Brimson MN 55602
Phone: 218-848-2885
FAX: 218-848-2433
Email: pberger@brimson.com

Paul Meysembourg
Principal Laboratory Technician
Natural Resources Research Institute

University of Minnesota
5013 Miller Trunk Highway
Duluth MN 55811
Phone: 218-720-4275
FAX: 218-720-4219
Email: paul@sparkie.nrri.umn.edu

James Sales
Principal Laboratory Technician
Natural Resources Research Institute
University of Minnesota
5013 Miller Trunk Highway
Duluth MN 55811
Phone: 218-720-4275
FAX: 218-720-4219
Email: jsales@sparkie.nrri.umn.edu

Carol Johnston
GIS Administrator
Natural Resources Research Institute
University of Minnesota
5013 Miller Trunk Highway
Duluth MN 55811
Phone: 218-720-4269
FAX: 218-720-4219
Email: cjohnsto@sparkie.nrri.umn.edu

Michael J. McCullagh

Quality, Visualization, and Use of Terrain Models in Physical System Modelling

---

Abstract

Terrain models require good visualisation systems to be useful in many physical system applications. The standard facilities found in many GIS modelling packages have been rapidly overtaken in terms of flexibility and photorealism by those increasingly available in true 3D modelling systems. These systems have become cheaper and ubiquitous on a wide variety of platforms, making the creation of detailed landscape views a possibility for all researchers.

Not only visualisation is required: animation is also a very necessary component to show dynamic aspects of geomorphological models. Once again the necessary systems are now present to generate animations reasonable easily, but quality real time animation is still really the personal province of rather expensive computer systems. Even so, a standard PC can now generate and play back such animations - given time! This paper looks at how these objectives may be achieved, and presents two movies of physical systems to illustrate the concepts involved.

Introduction

The accurate modelling of geographical terrain and similar rather ill behaved surfaces is one of the more difficult problems in surface modelling as it is very awkward to use simple or even complex mathematical functions which will perform reliably in all situations (Watson, 1992). More recent approaches using fractal methods have different but equally severe problems.

The methods used to develop models from digitally controlled stereoscopic air photographs are necessarily different from those used to generate terrain from field survey or even from pre-existing paper maps (Lam, 1983; Tobler, 1985; McCullagh, 1988; Weibel and Heller, 1991). This is because: (a) the data collection systems are wildly different, (b) the data quantities available from the source document are effectively unlimited and resamplable for the air photograph sitting on a stereo plotter, but cannot be enhanced without further survey for paper map or field survey source, and (c) most importantly, the model requirements and accuracies are likely to be very different for the two applications.

Once a method appropriate to the problem in hand has been determined, the platform and software must be chosen. The exact selection will depend on many factors, but there is an increasingly large range of systems available on low end computing machines that, in terms of user friendliness if not always in terms of complete functionality, are satisfactory substitutes for much larger and more expensive systems. The cheaper systems found on PCs have an advantage as well: they usually have support provided to allow the migration of the terrain models they have generated into a plethora of excellent viewing software. This includes pathways to full photo realistic images, and animation systems.

Sources of Data

The wise terrain modeller always looks round for already digitised data - or complete models - before reaching for a paper map and manual digitising table. There are a growing number of sources for digitised data, of varying scale quality and reliability, available either by mail or Internet, and either free or outrageously expensive. In all cases the source of the data should be investigated to try to get some idea of its authenticity and expected accuracy. Some data sets have been deliberately degraded before entering the public domain, and may contain considerable induced inaccuracy.

If one wants a model of part of the USA the data is usually available either at media production cost or free from such sources as the USGS node of the National Geospatial Data Clearinghouse (on the World Wide Web at www.nsdi.usgs.gov), the Eros Data Centre (edcwww.cr.usgs.gov), and many other US government organisations. Ready built models are available for quadrangle data at various scales, as are contour line and other data sets. Many organisations dangle pages on the WWW. A good starting point for a list to ones worth searching, for all countries of the world, is held on the GIS server at the University of Edinburgh (www.geo.ed.ac.uk).

In Europe the situation is complicated by the fact that most governments do not place their topographic survey data in the public domain, but maintain copyright and charge "cost retrieval economic prices" for it despite their citizens having already paid for it through taxes. This means that their WWW pages tend not to hold data but order forms! Charging does have its good points. For instance the Ordnance Survey (www.open.gov.uk/ordsurv) has good coverage in terms of both models and digitised data available. Special arrangements have been made for education and researchers access to OS digital data - but so far only for seven very limited areas of the country. These can almost be guaranteed never to coincide with one's research interest areas. Hopefully this situation may change soon.

Data cost and availability often depends on scale. At scales greater than 1:1M digitised data becomes far cheaper and more easily available. The Digital Chart of the World, a vector product originally produced for the Defense Mapping Agency in the USA, is available on a suite of 4 CD-roms for about $200 from Chadwick-Healey. The set contains vector digitising of the ONC 1:1M map series. It is not very useful for terrain modelling on other than a regional scale, and then only in areas of reasonable relief as the contours - digitised in feet - are never closer then 250ft, and have a default interval of 1000ft! The CDs do contain, nevertheless, a full culture and administrative set in addition to the contours and provide a very valuable global resource. Reading the files on the CD is not easy as the format of the data is quite convoluted. This has enabled resellers to offer the same set imported into other proprietary GIS and mapping systems for more like $2000. Raster data is also commonly available at regional scales. Gridded height data from Spot and other newer satellites is also a good source of terrain modelling - if the price is right, if one is dealing with multi kilometre square areas rather than small site investigations, if one can afford the cost, and if one does not mind vertical inaccuracies in the final model of possibly at least 10m. Positional accuracy may well be worse. At global scales there are a number of 5 minute and better altitude raster data where altitude (and depth) models are readily available. The World Data Centre run by NOAA at NCAR in Boulder (USA, www.ngdc.noaa.gov) can provide these at little more than media cost.

Modelling Resources

There are a wide range of modelling systems available on a diverse selection of computer platforms. This discussion merely points out a few of these as examples of their types. Table 1 gives an approximate idea of the source, market, and platform, of a few modelling packages. In some cases they are stand alone systems, and in others they are part of a much larger and more comprehensive package of GIS or CAD facilities.

PC software is available either for normal purchase, or as shareware. Panacea is an example of the former but is in the DOS world at the moment. The archive at www.micros.hensa.ac.uk contains several shareware packages of which Landscape Explorer, a Windows based product is a good example. Its capabilities are limited in the size of data sets that it can hold, but its facilities for import, export, contour tracing from bit images, and interpolation are really quite sophisticated. As with many PC packages there is the common (and correct) assumption that developers of different wheels can be used interchangeably on the modelling wagon via all the usual interface formats. Thus visualisation may be the strong point of one package (such as Vista Pro), and modelling of another. Software for the PC can now be cheap enough that many systems can be used and put to work on those aspects in which they excel.

Table 1: Sample modelling systems on a range of platforms

	Model Name	**Authorship**	**Address & Availability**
PC Systems	Panacea	Siren Systems	michael.mccullagh @nottingham.ac.uk
	VistaPro	Virtual Reality Labs Inc	USA tel: 01-800-829-8754
	Landscape Explorer	WoolleySoft	100332.2104@compuserve.com, or micros.hensa.ac.uk
PC Systems & Work Stations	DTM/W/G	Intergraph	www.intergraph.com
	Microstation	Bentley	www.bentley.com
	Arc/Info	ESRI	www.esri.com
	Moss	Moss Systems Ltd	Barclays House, 51 Bishopric, Horsham, RH12 1QJ
	DtmCreate	LaserScan	www.lsl.com

The other attractive feature about PC modellers is that they tend to have genuinely friendly user interfaces. This often does not yet apply fully to systems that are to be found on both PCS and work stations, or on work stations alone. They all tend to be stand alone packages which are often quite difficult to interface to each other. Costs also tend to be high. Often the systems can deal with larger data sets, and sometimes with a greater sophistication. Moss is an example of a civil engineering CAD system that grew, and now has a good grip on the highway engineering market.

Numbers of vendors have approached terrain modelling along the route of GIS and remote sensing. The TIN and GRID modules in Arc/Info are probably the present GIS market leaders, partly because ESRI has such a firm grip on the market through a huge installed user base, and partly because it has quite good modelling facilities. Intergraph and Bentley, who started in CAD and then moved into GIS and remote sensing, have parted company but offer very similar systems, though sometimes on rather specialised hardware. The success of Bentley's port to the Windows NT operating system and possibly Windows 95 will be interesting to watch.

The advantages of the work station over the PC have traditionally been those of speed and being able to handle "real" problems. This is still so, but only to a very limited extent. Most modellers operating in the Windows environment can handle large data sets of more than a few hundred thousand points successfully and reasonably easily merely by adding (quite cheap) memory. Disk space is no longer a limitation as disk prices have crashed making the cost of a 2GB disk the same as that of a 540MB disk last year. Speeds have risen along with increasing algorithm and graphic interface sophistication. The cost of a PC surface modelling system at anything from free to shareware to a maximum of a few hundred pounds should be compared with that of work stations which tend to hover around the low thousands of pounds.

Table 2: Sample visualisation and animation software

Class	**Name**	**Functionality**	**Cost**	**Availability**
PC: DOS	Idrisi	Drape Overlays	100	Clarke University
PC: Windows	VistaPro	Render & Animation	85	USA tel: 01-800-829-8754
PC: Windows	ENVI	Render & Fly Through	2000+	support@floating.demon.co.uk
PC Shareware: DOS	PolyRay	Ray Trace & Animation	25	xander@mitre.org & micros.hensa.ac.uk
PC Shareware: Windows	PoVCAD	CSG Modeller	25	72114.2060@compuserve.com
PC: Windows	Animator Studio	CSG Modeller & Animator	400	AutoDesk Ltd, tel: 01483-303322
PC: Windows	3D-Studio	CSG, Ray Trace & Animation	2000	AutoDesk Ltd, tel: 01483-303322
PC: Windows & Windows 95	Dream 3D	Ray Trace & Animation	300	Corel 6
PC: Windows	Caligari TrueSpace	Ray Trace & Animation	400	Roderick Manhattan Group
PC Shareware: Windows	GoldWave	Sound Capture, Player, & Editor		chris3@garfield.cs.mun.ca & micros.hensa.ac.uk
PC Shareware: DOS	DTA	Dave's *.TGA Animator . . .	25	76546.1321@compuserve & micros.hensa.ac.uk
PC Freeware: Windows	AAPlay	plays *.fli and *.flc animations	0	AutoDesk Ltd, tel: 01483-303322

PC Shareware: DOS	CMPEG	Create MPEG movies without sound	0	stefan@lis.e-technik.tu-muenchen.de & micros.hensa.ac.uk
PC Shareware: Windows	VMPEG (lite version)	Plays MPEG movies	0	stefan@lis.e-technik.tu-muenchen.de & micros.hensa.ac.uk
Work Station	Arc/Info	Drape Overlays	500+	www.esri.com
Work Station	Performer	Real Time graphics	1000	but you need an SGI machine, preferably a top range Onyx!

The Use of Models

Software designers often do not see beyond the creation of a model to where it will be used by others in many different areas. The range of applications is vast (see Moore, Grayson, and Ladson, 1991, for a comprehensive review), from ridiculous but very effective backdrops for animated fantasy to movies to sublime research measurement on damp hill slopes. Scale, source of data, method of construction, and need all vary but have one major requirement in common: they all require some form of visualisation and possibly animation to allow interpretation and understanding to acheive success in the task to which the terrain model has been put. It has been long established practice to represent models in the form of contour maps, probably as a matter of reassurance to the user rather than as a useful device. Similarly isometric and wire frame diagrams were very popular displays, sometimes overlaid with manually added information, and occasionally in stereo (McCullagh and Sampson, 1972).

Fortunately matters have progressed (McLaren and Kennie, 1989), and computer graphics (Earnshaw and Wiseman, 1992), and particularly the photo-realism school of graphics (Watts and Watts, 1992), have now provided the tools to appreciate modelled terrain as a landscape draped with overlying imagery and possibly with artificial constructs representing the user's research findings or other application. The impact of viewing landscape, correctly proportioned, complete with any vegetative and human culture, and analytical results, allows a researcher not only to present findings in papers and talks in an easily assimilable fashion (Hearnshaw and Unwin, 1994), but also provides the key to using that powerfully analytic but organic device, the eyeball-brain computing machine. If in addition that scene is animated to provide not just movement of the view point location but to allow the "furniture" to move as well, much more may be gained in visual and analytic terms. For example the sun could be moved realistically and insolation characteristics calculated exactly, or the results of fluvial, hydrological, oceanographic, or erosional simulations can be displayed in "real" lapse-time. Use can be made of the growing variety of modelling tools. This includes not just the well established constructive solid geometry CAD tools that generate such robotic views of the world, but the newer graphics constructs such as fractals for surface roughness, and procedural animations such as particle systems for droplet flows and analytical systems for wave and other motions.

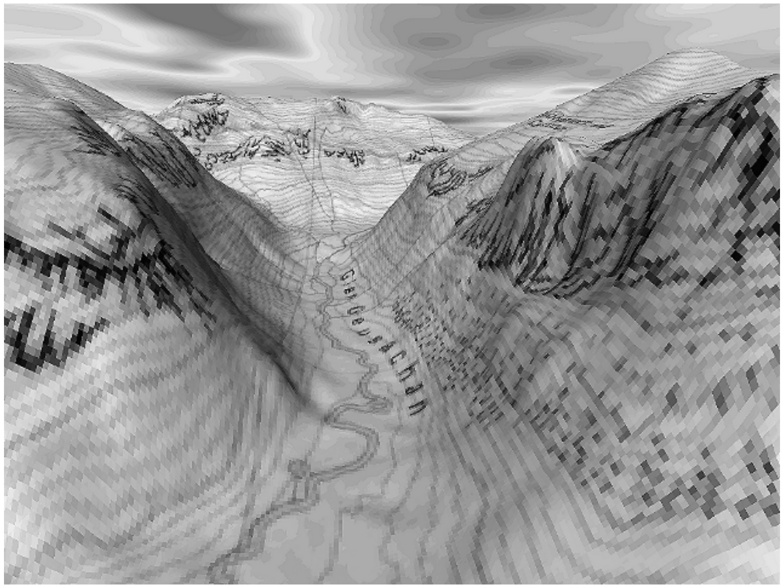


Figure 1: PolyRay Generated View of Valley in Cairngorms

A reasonable question is how much does all this cost and, for the maturer and wiser researcher, a rather more cogent one is how long does it take to create the views and animations? The answers, as for generating basic terrain models, vary from nothing to everything. Table 2 may help to put the problem of costs and performance into perspective. Costs can be very low, or even free using freeware or shareware. The products in this category are very good modellers indeed, but have a fairly hostile general user interface in the sense that the control of any scene has frequently to be written in a "C" like format. This is actually quite straightforward once one gets used to it. The example of PolyRay use given in Table 3 generates the basic valley view shown in Figure 1. A very simple script generates the entire view.

Table 3: Command script in PolyRay format to generate valley view

```
// Polyray Valley View File: Michael McCullagh
include "d:\polyray\dat\colors.inc"
include "plane.inc"
// Set up the camera
viewpoint { aspect 1.5 from < 0, 3.5, -8> at <1, 0, -1.5> up <0,1,0> angle 40 }
// Set up background color & lights
background sky_blue
light <50, 300, -500>
define os_map image("cairnmap.tga")
define os_shading
texture {
 special surface {color planar_imagemap(os_map,P,1) ambient 0.2 diffuse 0.8 specular 0.2}
 }
define cloudy_sky
texture {
 special surface { ambient 0.9 diffuse 0 specular 0
 color color_map(
 [0.0, 0.6, <0.4, 0.4, 0.4>, <1, 1, 1>]
 [0.6, 0.8, <1, 1, 1>, <0.196078, 0.6, 0.8>]
 [0.8, 1.0, <0.196078, 0.6, 0.8>, <0.196078, 0.6, 0.8>])[noise(3*P, 3)]
 }
 }
// Now set up the cloudy sky as a very large sphere
object { sphere <0, 0, 0>, 1000 scale <1, 0.01, 1> cloudy_sky { scale <50, 50, 50> }
 shading_flags 0 rotate <5, 0, 0>
 }
define rotation 60
object { smooth_height_field "valley.tga" os_shading
 scale <12, 0.01, 8> translate <-6, 0, -4> rotate <0, rotation, 0>
```

}

PolyRay is only one example of a large number of ray-tracers and animator programs available from the Internet. The www.micros.hensa.ac.uk address has more, and many others are referred to on the various graphics bulletin boards. More sophisticated Windows user interfaces and possibly faster renderers tend to cost money. The 3D-Studio software is very good, but is now threatened by a number of upcoming commercial rivals such as Caligari, and most recently, Corel's Dream 3D newly introduced as part of the Corel 6 package.

Animation Processes

The time to create single frame reasonably high resolution photo-realistic views on any platform tends to be measured in minutes at least. Time to create a fully ray traced animation sequence of perhaps 30 seconds varies from many days on your desktop PC, to only a few days on a common or garden work station, to almost real time on a purpose built Silicon Graphics Onyx multi-processor 2GB main memory and gobs of disk machine costing at least $200,000. Even on the Onyx using Performer software, the picture had better not be too complicated or the frame rate drops below a minimum acceptable 15fps. Even the 15fps limit in real time animation for this state of the art machine does not include full photo-realism ray-traced standards, but is good enough to fool the eye most of the time on simple scenes. There is a lot of scope for both speeding up the process and cost reduction. But if you have a PC on your desktop, why not put it to work over the night time and weekends and create the real thing without the mega expense involved in high priced work stations?

The plethora of digital video formats and sound tracks for recording digital animations is still a major problem, and will continue to be for some time until the standards finally settle down. The difficulty is that 30 seconds of animation at 25fps using the low resolution images (320 by 200) considered just good enough to show on NTSC (USA standard) television occupies 144MB of disk space in uncompressed form. This is without considering that essential feature of all animations - the soundtrack - which probably adds about 2MB to the total. Two problems result: that of transferring the video data sufficiently fast from disk to screen so no Chaplinesque results occur, and that of storing a sensible quantity of video data (say an hour) in a sensible amount of disk space. One hour might require a ridiculous 20GB of disk space.

The solution is compression of both sound and vision using either software or hardware appoaches. Individual images can be compressed either without loss using GIF, TIF, BMP and other formats, or with an "acceptable" loss using the more complex encoding provided by JPG (Joint Picture Expert Group) and similar formats. This typically leads to compression factors of 75% or more for full colour images. Video compression can go one stage further by compressing not just within each frame but by looking at the differences between successive frames, and only encoding areas that change. This can give compression of over well over 95% for many video sequences. Number of systems exist, but they tend to be dominated now by the older AutoDesk Animator formats (FLI and FLC), and the newer MPG (Motion Picture Expert Group) format that handles both sound and vision and achieves compression levels about 3 times higher than FLI/FLC. The DTA software listed in Table 2 is very effective at providing all forms of Animator output with excellent colour. The CMPEG program is rather less effective in terms of colour, but generates reasonable MPG sequences. The penalty that must be paid for "small" file sizes measured in only some hundreds of megabytes per hour of animation is not only some "acceptable" loss in picture quality, but also a substantial (though only measured in seconds per frame) software encoding process. The decoding can be performed by software (eg Vmpeg in Table 2), or by hardware boards which are fast, reasonably cheap, and widely advertised in the PC magazines. Software decoding on Pentium 100 PCs can reach over 25fps for "standard" 320 by 200 images. Hardware solutions can not only perform consistently in real time, but often come with encoding chips as well as decoding facilities and hence can be used for MPG movie generation as well as playback.

Animation Application

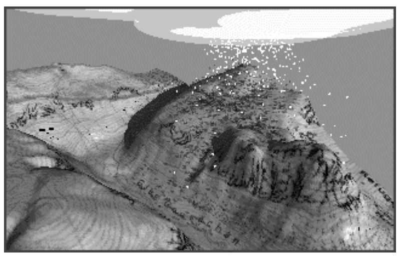


Figure 2: Storm Conditions in the Cairngorms

Animation provides the dynamic medium essential to comprehending complex spatial interaction models. Figure 2 shows a frame from a 30 second movie sequence. It employs a form of procedural modelling using particles. During the animation process particles can be created that effectivly have a life of their own. They can be made to obey physical laws of gravity, to interact with each other, and with the "scenery". In this case the particles considered are water droplets, but could for instance have been cars on a motorway. Particle systems and the development of analytical modelling of materials in computer graphics has made possible complex modelling of movement, not just of individual objects such as water droplets but also of connected particles such as skin, cloth, or water surfaces that respond to some form of wavelike motion.

The frame in Figure 2 is taken from an animation that attempts to show the possibilities for introducing particle systems into representations of erosional processes on landscapes. Hail, falling from the cloud at the top of picture, changes to water droplets on contact with the terrain surface. The hail falls in a semi turbulent fashion and lands randomly on the landscape. The droplets obey the laws of gravity before and after their conversion to water droplets. These then roll over the landscape, following the natural drainage channels in the model at a speed controlled by gravity and the local terrain. Ponding occurs frequently in low gradient areas, and waterfalls cascade off the corrie side. This is not intended to be an overly serious simulation of the overland flow of water, but is intended to show how animation can be used to demnostrate the products of simulation in a wide variety of environments. Similar animations could be made of flooding and tidal processes, wave action, and erosion.

Copies of the full animation sequence in FLI format, including a sound track in WAV format may be available (about 10MB in total) from this site or on the final CD.

Conclusion

Terrain modelling has come of age. There are many methods of creating models now available, many suited for a particular problem rather than a panacea for all surface generation. Models are now available for most parts of the world, though at varying levels of accuracy, scale, and cost. Even so there still seems to be a need for detailed models of specific small investigation site areas, and for good specialist models at a regional or world scale. Satellites, digital air photographs and GPS are rapidly filling some of the requirements, but owing to cost there is still a high demand for "home-built" models customised for particular purposes. Terrain models are now generated and used routinely in areas varying from hydrology to weather forecasting. The next step is realistic visualisation of those models and simulations, and animation of the visualisations to demonstrate the dynamic nature of many systems. Visualisation is still a challenge, although many of the tools are now in place, and fully (and freely in some cases) available to those who want to use them. Animation is in its infancy, as evidenced by the rapidly changing computer video standards being proposed, but already has tools available via the Internet. Let's go and use them!

REFERENCES

Earnshaw R.A. & Wiseman N., 1992, *An Introductory Guide to Scientific Visualisation*, Springer-Verlag, 156p.

Hearnshaw H.M. & Unwin D.J., 1994, eds, *Visualisation in Geographical Information Systems*, Wiley, 243p.

Hutchinson M.F., 1989, A New Procedure for Gridding Elevation and Stream Line Data with Automatic Removal of Spurious Pits, *Journal of Hydrology*, 106, 211-232.

Lam N.S., 1983, Spatial Interpolation Methods: A Review, *American Cartographer*, 10 (2), 129-149.

McCullagh M.J. & Sampson R.J., 1972, User Desires and Graphic Capability in the Academic Environment, *Cartographic Journal*, 9, 109-122.

McCullagh M.J., 1981, Creation of Smooth Contours Over Irregularly Distributed Data Using Local Surface Patches, Geographical Analysis, 13 (1), 51-63.

McCullagh M.J., 1988, Terrain and Surface Modelling Systems: Theory and Practice, *Photogrammetric Record*, 12 (72), 747-779.

McLaren A. & Kennie T.J.M., 1989, Visualisation of Digital Terrain Models: Techniques and Applications, *in* Raper J., *Three Dimensional Applications in Geographic Information Systems*, Taylor & Francis, 79-98.

Moore, I.D., Grayson R.B., & Ladson A.R., 1991, Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications, *in* Beven K.J. & Moore I.D., *Terrain Analysis and Distributed Modelling in Hydrology*, Wiley, 7-34.

Petrie G., & Kennie T.J.M., eds, 1990, *Terrain Modelling in Surveying and Civil Engineering*, Whittles Publishing, 351p.

Samet H., 1990, *The Design and Analysis of Spatial Data Structures*, Addison-Wesley, p.

Shepard D., 1969, A Two-Dimensional Interpolation Function for Computer Mapping of Irregularly Spaced Data, Harvard Papers in Theoretical Geography, Harvard Univerity, Paper 15, 20p.

Tempfli, 1986, Composite / Progressive Sampling, A Program Package for Computer Supported Collection of DTM data, *ACSM-ASP Convention*, Washington DC, 9p.

Tobler W.R., 1979, Lattice Tuning, *Geographical Analysis*, 11 (1), 36-44.

Tobler W.R., 1985, Smooth Multidimensional Interpolation, *Geographical Analysis*, 17 (3), 251-257.

Watson, D.F. , 1992, Contouring: *A Guide to the Analysis and Display of Spatial Data*, Pergamon, 321p.

Watt A., & Watt M., 1992, *Advanced Animation and Rendering Techniques: Theory and Practice*, Addison-Wesley, 455p.

Weibel R., & Heller, M., 1991, Digital Terrain Modelling, *in* Maguire D.J., Goodchild M.F. & Rhind D., *Geographical Information Systems*, Vol 1: Principles, Longman, 269-97.

Wood J.D. & Fisher P.F., 1993, Assessing Interpolation Accuracy in Elevation Models, *IEEE Computer Graphics and Applications*, 13 (2), 48-56.

The future of spatial modeling for understanding and predicting landscape transformations

Robert Costanza

Abstract:
Understanding and modeling the spatial patterns of landscape changes over time at several different scales is critical to effective environmental management. In recognition of this, the US EPA has recently begun movement away from their traditional "media-based" approach to environmental management and towards a "place-based" approach. To operationalize this approach, we need to develop a deeper understanding of the complex spatial and temporal linkages between ecological and economic systems on the landscape, and to use that understanding to develop effective and adaptive policies. This will require new methods that are comprehensive, adaptive, integrative, multiscale, pluralistic, and which acknowledge the huge uncertainties involved. Landscape modeling studies at local, regional, and global scales require the integration of natural and social sciences and the development of a common framework for understanding linked ecological economic systems.
Several case studies of regional dynamic landscape models are first presented to set the context. Key areas for future development include: (1) using the modeling process to build a broad consensus among stakeholders; (2) developing sufficient data bases of historical landscape changes; (3) linking process models from several disciplines with geographic data at several different time and space scales; and (4) understanding the relationship between resolution and predictability in landscape modeling.

Contact: Dr. Robert Costanza
Professor and Director, University of Maryland Institute for Ecological Economics
Center for Environmental and Estuarine Studies (CEES)
University of Maryland
Box 38, Solomons, MD 20688-0038
Phone: 410-326-7263
Fax: (410) 326-7354
email: Costza@cbl.cees.edu

Dennis S. Ojima and William J. Parton

Integrated Approach to Land Use Analyses

ABSTRACT

The recognition that growing human population levels with the associated vast consumption of the earth's natural resources is placing extraordinary pressure on the biological and chemical systems of the planet has become increasingly evident. Human activities are significant drivers of environmental changes and the analysis of the ecosystem response to changes in these human-derived environmental factors are critical to our understanding of current and future changes in our environment. Building on our understanding of ecosystems in response to environmental changes, we are now capable to integrate more directly the human dimension into our analytical framework of a human-ecological system. A more explicit treatment of the human system in conjunction with the ecosystem properties is needed. With increasing modifications of our natural landscapes, what choices will be made by resource and land managers become increasingly important to the long-term sustainable use of our natural resources and to human welfare. These decisions are influenced by economic, social and political forces. In addition, changes in ecosystems taking place in less developed regions of the world, such as in the Asian steppe, are being modified by similar forces and are also influenced by the technological institutions available to them. So the ability to evaluate the effect of changing ecosystem properties on the ability of humans to utilize various natural resources and to determine the effect of altering human activities on the availabilty of natural resources is needed to better assess the overall impact of these changes on sustainable human-ecosystem dynamics. This information will be useful to policy makers and local resource use managers, alike. This paper outlines the analytical framework of the assessment and illustrates how changes in land use affect environmental properties, such as system level carbon exchange.

INTRODUCTION

It is well-known that global scale changes in climate will affect land use patterns, and in turn, land use changes are likely to be an important factor in affecting global change by altering terrestrial carbon uptake/emission (Houghton et al. 1990). At local to regional scales, land management practices, such as fire, grazing, tillage practices and conversion to and from arable lands associated with agriculture, forestry, rangeland, and urbanization, affect ecosystem composition, cycling of nutrients, and distribution of organic matter (Houghton et al. 1983, Turner and Meyer, 1994; Schimel et al. 1991 and Ojima et al. 1994a,b, Parton et al. 1995). At broader scales, population growth and rising per capita incomes will continue to exert pressure on terrestrial ecosystems in the form of rising demands for food, fuel, fiber, water, and other resources. These extractive pressures, in many cases, override natural controls of the global environment and accelerate rates of global change to unprecedented levels (Bolin et al. 1986, WCED 1987). An analytical understanding of the complex interactions of climate and human activities in a region is a challenging but necessary task to accomplish in order to predict credibly the outcome of different management schemes and environmental changes integrated across environmental, economic, social, and political systems, referred to in this proposal as the "human-environmental system".

The influence of land use change on terrestrial carbon dynamics constitutes a core research focus of the International Geosphere-Biosphere Program (IGBP) (IGBP Report 21, [Steffen et al. 1992] and 24 [Turner et al. 1993]). Land use decisions are driven by mostly demographic and economic considerations at broad scales and by a medley of social, economic and environmental considerations at regional and local scales. Turner and Meyer (1991) recognized that the driving forces of land use change are different at different spatial scales, though some forces cut across all scales, such as population change. We agree, and argue that a hierarchical or spatially nested analytical framework is needed in order to understand how driving forces of land use and climate changes transform land covers at local to regional scales (Figure 1, from W.E. Easterling, unpubl. man.). How we scale local responses to regional dynamics of agroecosystems and land use will take more research, however an analytical approach to evaluate the information passed from one level to another is a critical first step.

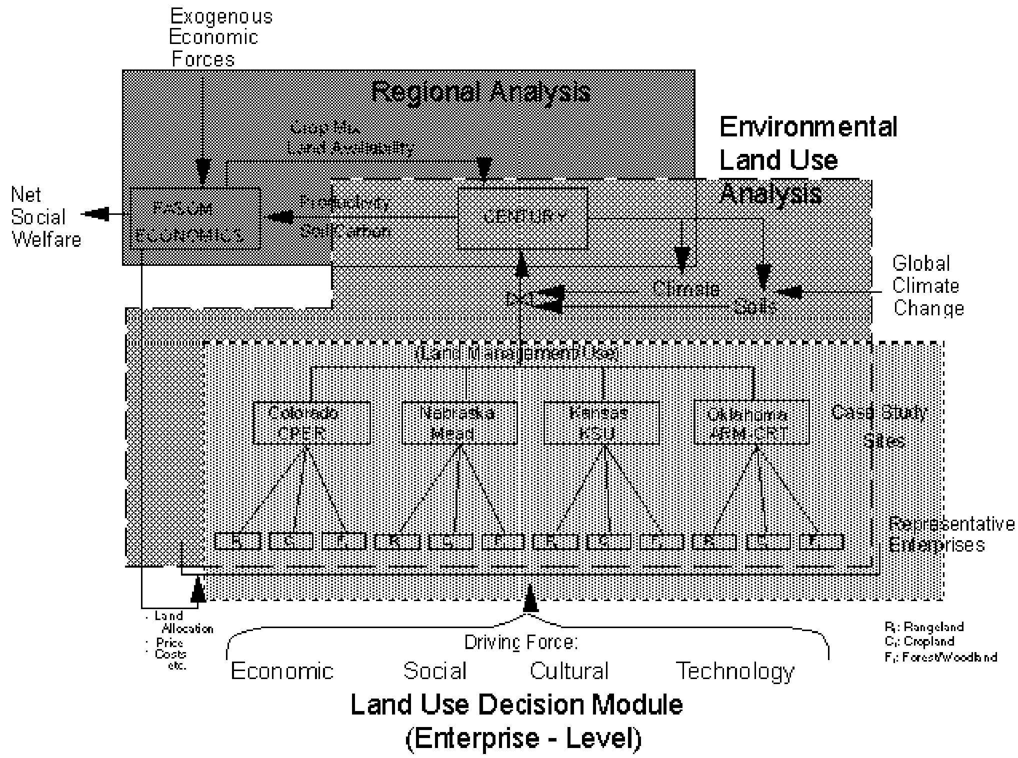


Figure 1. Analytical structure for assessment of land and climate change effects on agroecosystems of the US Great Plains.

In early studies of the effects of climate change on the agricultural sector (Rosenzweig 1985, Rosenzweig and Parry 1994), the results of crop yield models for representative areas were used to characterize climate impacts on yields at the regional and national levels. These studies tended to focus on changes in crop yields in specific locations and not on geographic shifts in cropping patterns. Crop losses projected in these studies were often quite severe, largely because they failed to account for the role of economic factors. More recent studies, using agricultural sector models, (Adams et al. 1990, 1995, Darwin et al. 1995) have shown how changes in climate can potentially alter cropping patterns in the United States due to climate-induced changes in the relative profitability of different crops. These studies indicate, by and large, that farmers can adapt to changes in climate by relying on market price signals to tell them how to adjust their crop mix and input use.

Our ability to predict ecosystem dynamics relative to climate or land use changes is dependent on the development of analytical tools to integrate our current understanding of how these ecosystems behave relative to human and environmental factors. The analysis of this information will need to incorporate the critical factors of the physical environment, including climate and soil factors, but also include the human factors related to land use practices (Figure 2). The dynamics of the agroecosystem will depend on the joint influence of the physical environment and the specific mix of land use practices implemented with a location. The land use decisions are controlled by a number of factors economics, policy, technological advancements, and socio- cultural factors.

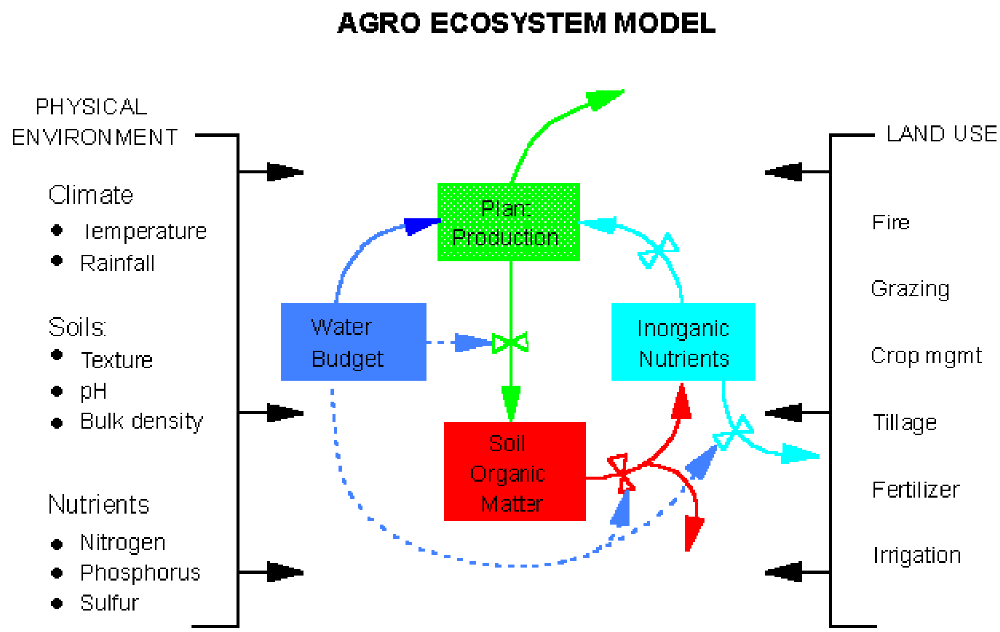


Figure 2. Simplified structure of the CENTURY agroecosystem model (Parton et al 1987, 1995) indicating the physical environmental controls and the set of land use management options implemented.

Assessment of long-term agricultural and ecological sustainability of these agroecosystems must be evaluated in relationship to variations in livestock abundance, cropping types, and other land uses are needed to assess the full range of ecological impacts of agricultural development. These predictions require a synthesis of the long-term effects of interactions between the human, environment, and the ecosystem components so that the impact of changes in one component can be evaluated relative to responses in other components of the overall system. For instance one would need to understand the influence of livestock on forage plant production and survival; the effects of human wood use on woody plant populations; the direct and indirect effects of livestock on soil structure and fertility; and agricultural use of water, soil, other resources, and technologies. The importance of past and current climate and land use cannot be overlooked in assessing the how these ecosystems has developed over the centuries and changes in the future relative to new policies, technological advances, economic conditions, and environmental constraints. A framework to simplify the complex interactions within and between various subsystems is provided using a modeling approach that includes all the major components and links them together in a spatially integrated fashion (Figure 3). At the core of this approach is an ecosystem-level process model, CENTURY (Figure 2, Parton et al, 1995) that incorporates the changes in external forcing factors, including climate and management on net primary productivity, nutrient and water availability, and carbon and nitrogen fluxes.

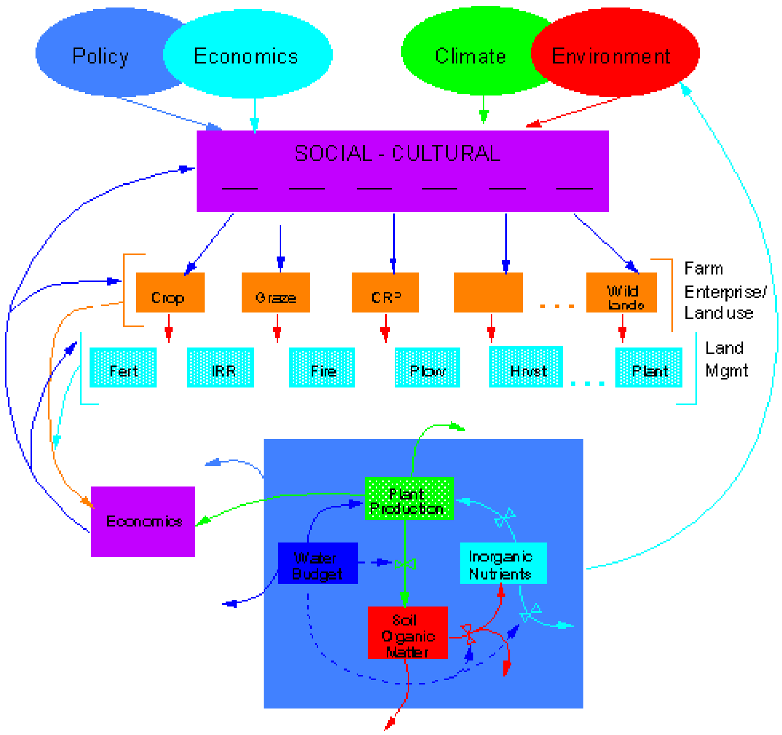


Figure 3. Linkage of social-cultural-economic factors influencing the land use decisions which modify agroecosystem processes.

Ecosystem Modeling

The CENTURY model can simulate specific land use systems within the region. Using different aggregation schemes for land use, soil, or climate factors, ecosystem dynamics can be evaluated relative to environmental and land use characteristics at the local to regional scale. Based on an examination of these factors we evaluated the influence of different land use management practices. Information needed include records of weather and plant productivity, land use history, and socio-economic trends. Many factors and decisions bear directly on land use, including what land units are available for grazing or cropping, soil properties, the severity of the climate, price of various input and output commodities, cultural and political constraints to the type of cropping system, or livestock system.

In recent analysis of assessing changes in land use management in the corn belt region of the United States, CENTURY simulations of improved land use practices resulted in a recovery of 47 to 79% of the initial soil C lost after approximately 50y of cropping (Figure 4, modified from Parton et al., 1995, Donigian et al. 1995). These analyses indicate that changes in land management practices can affect C storage in the soil without greatly affecting yield of corn in this region. The factors influencing crop rotation and crop selection also affected the level of C stored in this mesic region of the US (Donigian et al. 1995).

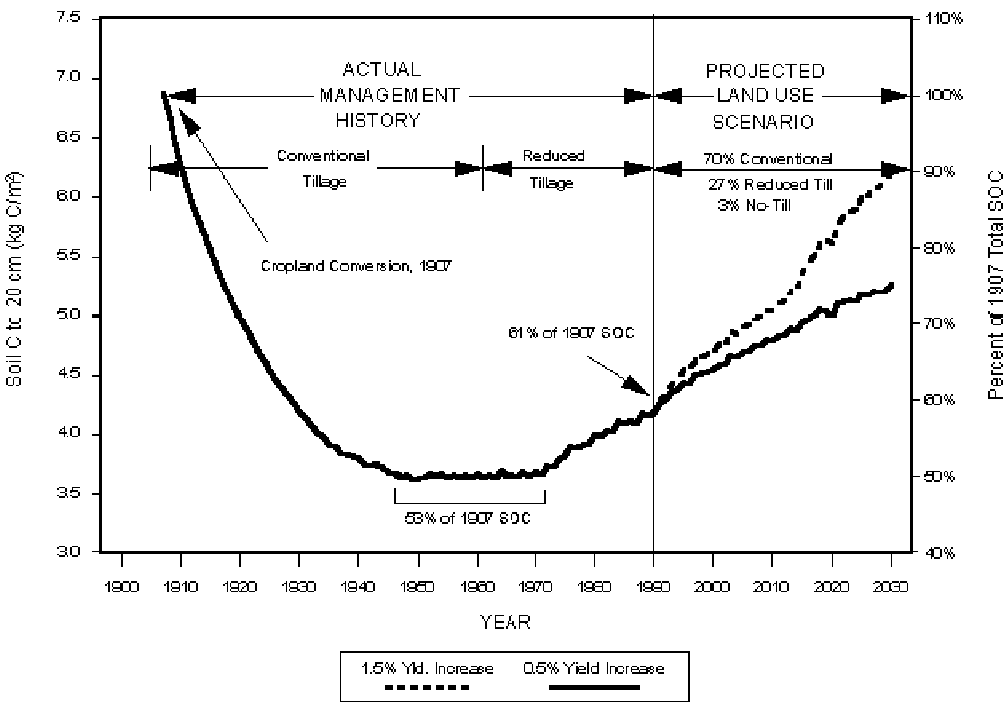


Figure 4. The effect of changing land management in the cornbelt of the United States on soil carbon levels in the surface 20 cm of soil.

Analysis of factors influencing land use decisions indicate that economic in certain situations, land use decisions are directly related to inherent properties of the physical environment, such as soil quality and moisture availability (Bohren and Riebsame, pers comm). Cultural and economic factors come into play as a set of factors that constrain these land use decisions (Riebsame et al 1994). Economics come into the analytical framework at the local scale of the farmer/rancher-level and at the regional level where global to regional scale prices and commodity information affects social and cultural decisions at this scale.

Technological advances in remote sensing (RS), geographic information systems (GIS) and ecological simulation modeling enable us to link information across a broad array of disciplines. The manner in which we will incorporate the case study results into a regional assessment will be based on these advancements in GIS and regional modeling of physical information (McKeown et al, this volume) and translating the social-economic information in this analysis (Figure 4). This analytical capability is needed to answer a variety of complex questions related to changing patterns of environmental and socio-political drivers in the Central US. Since land use change is a dynamic process, the integration of GIS with simulation techniques provides a way to examine their spatial and temporal characteristics and identify forces contributing to land use change. Analysis of the land cover data base developed by the EROS Data Center of the USGS indicates that cropland conversion tended to be on the richer soils of the Platte River drainage of eastern Colorado. Areas of lesser soil C tended to remain in rangeland. Using these relationship predictions of how projected changes in land use can be determined in spatially explicit manner. The utility and information content of GIS and RS data that incorporates the socio-political and economic factors with that of ecosystem structure and function will greatly increase our ability to interpret land use-ecosystem processes.

Incorporating information describing the historical changes in land use, the CENTURY model has been used to simulate the impact of cultivation in the Great Plains during the last 100 years. The model results were compared with soil organic matter data and county level crop yield data for that time period. The model correctly predicted the patterns of increasing crop yields and decreasing soil organic matter and soil N mineralization rates during that time period. The major inputs to the model include weather data and the historical patterns of cultivation practices, varieties for the major crops and amounts of inorganic fertilizer inputs. The increase in crop yields resulted from the use of higher yielding varieties, improved cultivation practices and increases in inorganic fertilizer levels.

Implementation of these land use practices in a spatial modeling and GIS scheme over larger areas will require a larger network of information flow. Monitoring over larger regions can be achieved through a hierarchically organized system where information flows upwards from local studies to provide assessments over larger spatial areas. Local information is integrated to understand region-wide response to environmental conditions (Figure 5). Spatial variations in landscape properties must be integrated with ecosystem processes to understand or predict the dynamics of the agroecosystems.

CONCLUSION

In this period of rapid environmental and human-system changes, an integrated approach to assessment of the ecological and human-system dynamics is fundamental to our understanding and assessment of ecosystem dynamics. Changes in land use and climate systems are contributing to rapid changes in how ecosystems behave. Ecosystem and socio-economic models are needed to investigate the differential causes and effects of land use changes. Spatially integrated data need to be linked to these models. These models also need to be developed within the context of a policy formulation decision support system, so that the results will be not only of scientific interest but of direct policy relevance.

Information on various levels of social institutions needs to be evaluated to identify the manner in which decisions are made that determine land-use systems under political units in the region. Models need to be able to reconstruct historical sequences of land uses and ecological integrity along a gradient of environmental factors. With this set of information and models, we will be able to evaluate various scenarios of land use and climatic changes. The papers presented in this unit illustrate the utility of an analysis framework that incorporates information and models for analysis of changes in environmental and socio-economic factors needed to understand the processes controlling changes in agroecosystems of the US relative to environmental and human-related changes.

ACKNOWLEDGMENTS

Support for research and assessment technology development is provided by NSF-Methods and Models for Integrated Assessment Program, NASA/Mission to Planet Earth Program, and NIH funding. We thank Becky Techau, Brian Newkirk, and Michelle Nelson for technical support at the Natural Resource Ecology Laboratory, Colorado State University.

REFERENCES

Adams, D.M., Alig, R., Callaway, J.M., McCarl, B.A. (1994) *Forest and Agricultural Sector Optimization Model: A Description*. RCG/Hagler, Bailly, Inc. PO Drawer O, Boulder Colorado 80306-1906. p54- 59.

Bolin, B., Doos, B.R., Jager, J. and Warrick, R.A. (1986) *The Greenhouse Effect, Climate Changes and Ecosystems. A Synthesis of Present Knowledge*. John Wiley and Sons, Chichester, London, 541 pages.

Darwin, R., M. Tsigas, M., Lewandrowski, J., Raneses, A. (1995) World Agriculture and Climate Change: Economic Adaptations. *ERS Agricultural Economic Report No. 703*. United States Department of Agriculture, Washington, D.C.

Donigian, A.S., Jr, PatwardhamA.S., Jackson, R.B. IV, Barnwell, Jr., T.O., Weinrich, K.B. and Rowell, A.L. 1995. Modeling the impacts of agricultural management practices on soil C in the Central US. P.121- 145. In R. Lal, J. Kimble, E. Levine, and B.A. Stewart (eds.), *Soil Management and Greenhouse Effect*. Advances in Soil Science. CRC Press. Boca Raton, FL.

Houghton, J.T., Jenkins, G.J. and Ephraums, J.J. (eds.) (1990) *Climate Change. The IPCC Scientific Assessment*. World Meteorlogical Organization (WMO), Cambridge University Press, Cambridge.

Ojima, D.S., Schimel, D.S., Parton, W.J. and Owensby, C. (1994) Short- and long-term effects of fire on N cycling in tallgrass prairie. *Biogeochemistry* 24:67-84.

Ojima, D.S., Galvin, K.A. and Turner II., B.L. (1994) The global impact of land-use change. *BioScience* 44:300-304.

Parton, W.J., Schimel, D.S., Cole, C.V., and Ojima, D.S. (1987) Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Science Society of America Journal* 5:1137-1179. **465**

Parton, W.J., Ojima, D.S., and Schimel, D.S. (1995) Models to evaluate soil organic matter storage and dynamics. In M.R. Carter and B.A. Stewart (eds.) *Structure and Organic Matter Storage in Agricultural Soils.* Advances in Soil Science. CRC Press. Boca Raton, FL.

Rosenzweig, C. (1985) Potential CO2-Induced effects on North American Wheat Producing Regions. *Climatic Change* 7:367-89.

Rosenzweig, C., and Parry, M. (1994) Potential Impact of Climate Change on World Food Supply. *Nature* 367: 133-138.

Schimel, D.S., Kittel, T.G.F. and Parton, W.J. (1991) Terrestrial biogeochemical cycles: global interactions with the atmosphere and hydrology. *Tellus* 43AB:188-203.

Turner, II, B.L. and Meyer, W.B. (1991) Land use and land cover in global environmental change: considerations for study. *ISSJ* 130:669-679.

Turner, II, B.L., Moss, R. and Skole, D. (1993) Relating land use and global land-cover change. Stockholm. International Geosphere-Biosphere Program. *IGBP Report #24/HDP Report #5*.

Turner, II, B.L. and Meyer, W.B. (1994) Global land-use and land-cover change: An overview. In: *Changes in Land Use and Land Cover: A Global Perspective*. W.B. Meyer and B.L. Turner II (eds.). Cambridge University Press, p. 3-10.

World Commission on Environmental and Development (WCED). (1987) *Our Common Future*. Oxford University Press, Oxford.

Dennis S. Ojima
Senior Research Scientist
Natural Resource Ecology Laboratory
Colorado State University
Phone 970 491 1976
Fax 970 491 1965
email dennis@nrel.colostate.edu

William J. Parton
Senior Research Scientist
Natural Resource Ecology Laboratory
Colorado State University
Phone 970 491 1987
Fax 970 491 1965
email billp@nrel.colostate.edu

Spatial resolution of crop models in the estimation of regional agroecological effects of climate change: how fine is fine enough?

William E. Easterling

Abstract:
The spatial resolution of climate change scenarios from general circulation models (GCMs) is too coarse (three to five hundred kilometers) to use in physiologically-based crop growth models to simulate economic yields accurately at a regional level. Crop simulation models are intended to simulate growth processes of a single plant or across a hectare. GCMs will never attain the resolution of crop simulation models and it is doubtful that crop models will ever mechanistically simulate crop growth at resolutions of hundreds of kilometers. Thus, there is a basic scale mismatch between climate models and crop models that has not been adequately addressed. Recently, however, meso-scale climate models have been successfully "nested" within GCMs in order to "down-scale" climate change predictions to levels that are useful in impact analysis.

The missing link in the continuum from crop models to climate models is the development of a scheme for "up-scaling" location-specific crop models to the resolution of meso-scale climate models (approximately 40-50km resolution). In this paper, different levels of resolution of climate and soils input data are examined within the Erosion Productivity Impact Calculator (EPIC) crop growth model in order to determine the spatial scale of such data at which there is maximum agreement between modeled and observed crop yields in the central Great Plains. Once that scale is determined, climate change impact simulations will be performed with scenarios from the Meso-Scale Model Version 4 (MM4) of the National Center for Atmospheric Research and compared with previous impact simulations using only GCM-scale climate changes.

Preliminary results indicate that spatial disaggregation of climate data from the GCM scale (320-550km) to approximately 120km greatly improves agreement between modeled and observed yields. Disaggregation of soils data does little to improve aggreement. Further disaggregation of climate data to 40km resolution (scale of the MM4) yields little marginal improvement in aggreement beyond that achieved with the 120km resolution. This work will provide climate data inputs to ongoing research on interactions of climate and land use change with agroecosystems being performed in collaboration with D. Ojima and W. Parton at the Natural Resources Ecology Laboratory at Colorado State University.

Contact: William E. Easterling
Department of Agricultural Meterology
237 L.W. Chase Hall
Univeristy of Nebraska-Lincoln
Lincoln, Nebraska 68583-0728
Phone: 402-472-6283
email: agme014@unlvm.unl.edu

Richard O. Flamm

Integrating spatial models into local land-use decision making

ABTRACT

In Florida, land-use decisions at the local level range from individuals operating watercraft to the designing of an infrastructure for urban development. These decisions, which individually are often considered relatively innocuous, accumulate to have measurable impacts on landscape structure and function. Landscape ecologists who seek to include the "human dimension" in their studies need to understand how humans use information to make decisions. For example, humans use relatively simple spatial models, typically maps of what's present, to make a multitude of land-use decisions. As modelers, we know that more complex manipulations of spatial data can provide additional knowledge about the system and its functions. Ecology, which has had tremendous success in model derivation, has had little impact on decision making at the local level, partly, because of the difficulty of introducing it's models into public arenas.

This paper discusses integrating ecological models into land-use decision making at the local level. Examples are presented of spatial models now being used in land-use decision making; a new, more complex model that has yet to be introduced; and scenarios in which model-based land-management alternatives for South Florida's natural system boundary are evaluated based on the landscape structure and function of that area. A computer-based modeling environment called the Land Use Change and Analysis System (LUCAS) is presented as an organizational framework for integrating ecological models into land-use decision making.

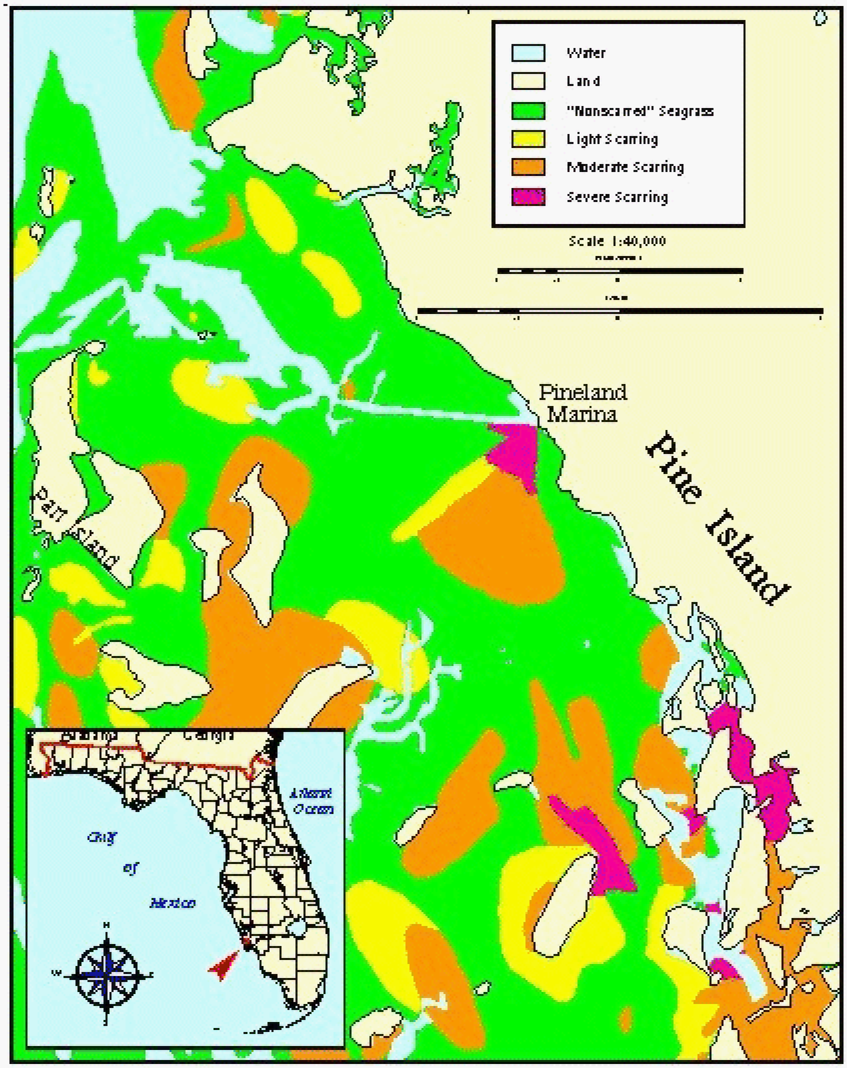
INTRODUCTION

In Florida, land-use decisions at the local level range from individuals operating watercraft to the designing of an infrastructure for urban development. These decisions, which individually are often considered relatively innocuous, accumulate to have measurable impacts on landscape structure and function. Landscape ecologists who seek to include the "human dimension" in their studies need to understand how humans use information to make decisions if they expect their models to influence future land use and ultimately have a positive impact landscape structure and function. Unfortunately, the ways humans make decisions is complicated and not easily modeled and is probably the major contributor to the error component of any predictive land-use model. Therefore, for several projects in Florida, we are examining land use decision making by applying spatial models of land use within sociological frameworks.

Three examples of integrating spatial models into local land-use decision making are presented: a spatial model now being used in local land-use decision making; a new, more complex model that has yet to be introduced; and scenarios in which model-based land-management alternatives for South Florida's natural system boundary are evaluated based on the landscape structure and function of that area. I conclude with a discussion of some issues that research teams will likely face during their attempts at introducing their spatial models into land-use decision making.

The most common spatial model used in local decision making is a map of what's present. Although not simple to construct, maps of what's present are probably the easiest models to integrate into land-use decision making. This is not very surprising given that they have been used for centuries, are well accepted, and although specifics of maps differ among cultures, overall, people's perception of what they mean is fairly uniform.

One large project that has been undertaken in Florida's Department of Environmental Protection is the mapping of seagrass scarring caused by boats (Sargent et al. 1995). Scarring is of significant environmental concern in Florida because of the continuous decline in seagrass coverage and the importance of seagrass as benthic habitat. In this project, the goals included (1) building a simple land-use model that informed the target users -- resource managers -- about land-use impacts and (2) influencing future land use in a positive way.



Detailed map of scarred seagrasses - Pine Island, Lee County

Fig. 1. Map of seagrass scarring in Pine Island, Lee County (From Sargent et al. 1995).

Maps of scarring were derived from aerial photos (Figure 1). Scarring was categorized as light, moderate, and severe. As maps of what's present, these models were easily understood, were widely distributed to resource managers, and have had a positive impact on local land use. They are being used on signs to inform boaters of the presence of seagrass, to determine where to place navigational markers directing boats away from seagrass beds, and to delineate closure zones so that seagrass beds can recover. The result has been that many seagrass beds are recovering, indicating beneficial changes in land use.

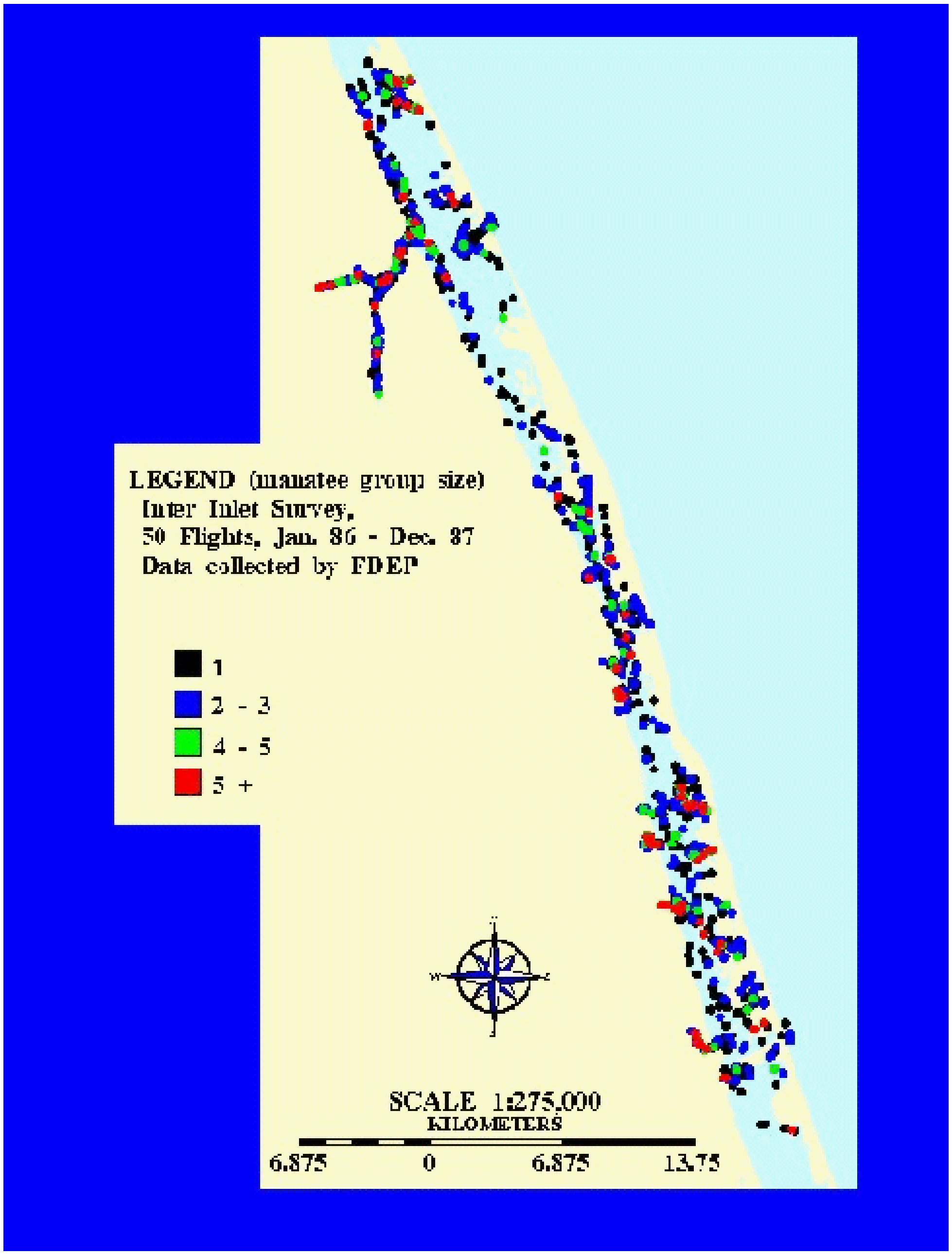


Fig. 2. Map of manatee-sighting data collected during aerial surveys along a route that ranged from Fort Pierce to above the Sebastian River, called the Inter-Inlet, along the west coast of Florida.

The second example, a more complex analytical model in the process of being introduced into land-use management, is associated with manatee protection. The manatee is an endangered herbivore that has been negatively impacted by a burgeoning human population along the coast of Florida. The dominant manatee mortality factor -- collisions with power boats -- is responsible for approximately 25% of their annual mortality. Thus, one important component of manatee protection is the designation of slow speed zones for boats in areas of high manatee abundance. Currently, these areas are delineated, in part, through visual inspection of aerial survey data plotted onto a map of Florida coastline (Figure 2). The problem with this approach is that the aerial survey maps can be difficult to interpret, which can cause readers to suggest differing delineations.

One task of the Marine Mammal program in the Florida Department of Environmental Protection was to devise a method to objectively delineate high-abundance areas from aerial survey point data. The method we developed and are experimenting with is a spatial filtering algorithm we call the fixed-area flexibly-shaped spatial filter.

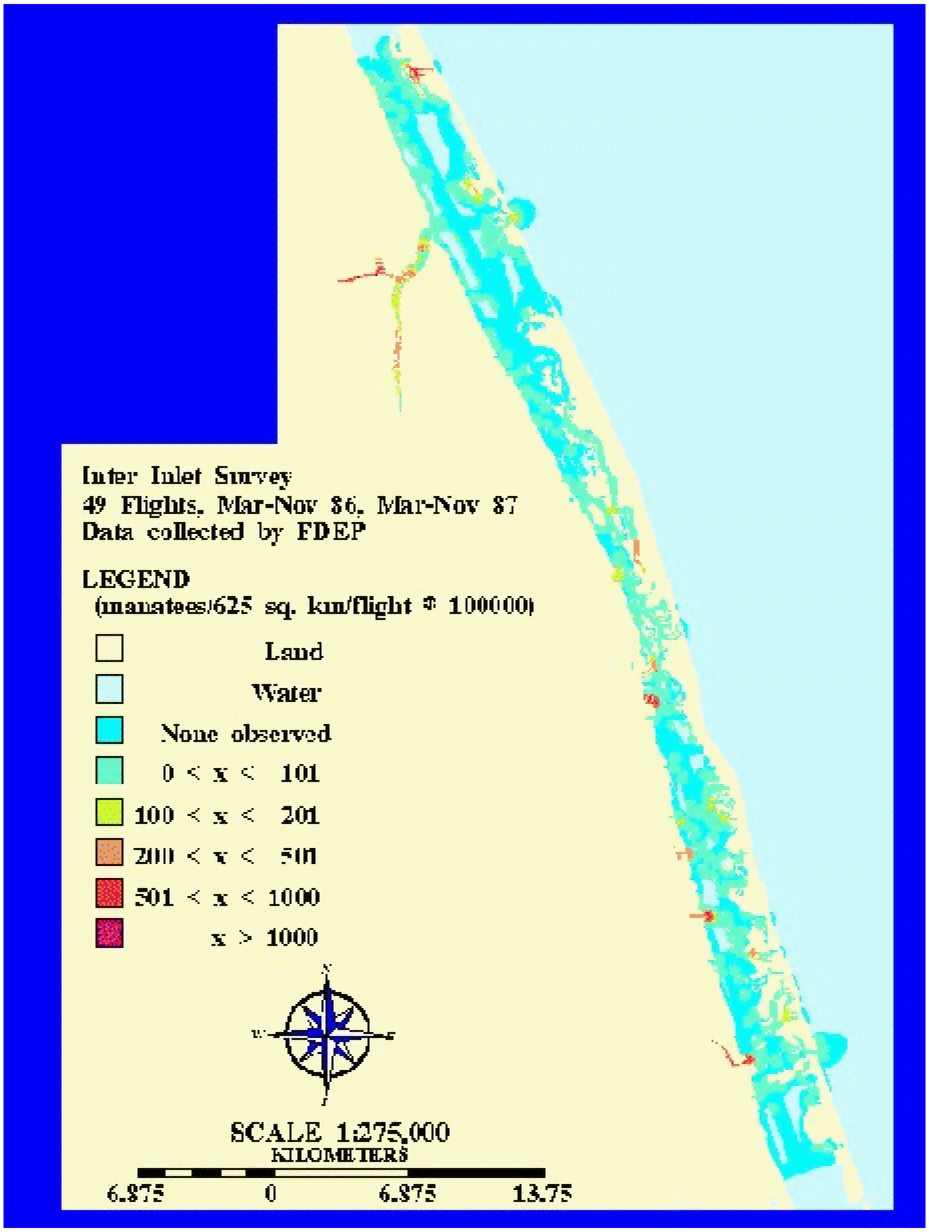


Fig. 3. Resu lts of applying fixed-area flexibly-shaped spatial filter to aerial survey point data collected along the Inter-Inlet survey route that ranged from Fort Pierce to above the Sebastian River along the east coast of Florida. Polygon size of the filter was 325 square meters.

Running the algorithm results in a contoured surface representing manatee abundance derived from aerial survey point data (Figure 3). The cell values in this map are interpreted as the likelihood of observing an animal while traveling along the aerial survey route. By applying rules in a GIS, such as "locate all areas with a manatee abundance value greater than 0.002 manatees/pixel," we can use these maps to help designate protection zones.

Using the spatial-filter model in decision making requires a different approach than that given in the seagrass scarring example because the target users are different and the model is less intuitive. The targeted audience for the seagrass-scarring model was primarily resource managers so the output of the model was catered to their uses. The spatial-filter model is targeted toward individuals or organizations that either make or are affected by policy. As such, it would be introduced in public forums such as local meetings and court challenges where the bases of decisions are subjected to close scrutiny. It is inevitable that stakeholders such as the members of the Marine Industry Association or of conservation organizations, and State and local officials will have various levels of mathematical or GIS expertise. Most of these people would probably be uncomfortable with decisions based on the model because they do not understand the underlying methodology. At present, this model is not used in manatee protection planning.

In an attempt to integrate this model into manatee protection planning, we adopted a strategy of educating policy makers and the public through an organization called the Manatee GIS Working Group. The working group consists of people interested in using GIS or understanding the role of GIS in manatee protection. Topics of interest to the Working Group include data-sharing issues, analysis techniques, and exploring decision making. The goal of the modelers is to use the manatee spatial-filter model to acquaint stakeholders in the working group with the filtering procedure and to facilitate the procedure's understanding and acceptance in public forums. We are just beginning to examine analytical techniques in the working group, so an evaluation of the group's effectiveness is premature. However, at least one stakeholder has demonstrated his committment to the process by hiring GIS expertise to represent them at the meetings.

The last example is a land-use project in its infancy called the Governor's Commission for a Sustainable South Florida. This project seeks to examine local land use and the influence of broad-scale policy on the landscape integrity of South Florida. Early phases of this program will involve simple manipulations of delineated "natural" areas. However, the project is expected to evolve into a more sophisticated integrated modeling system whereby users can evaluate alternative natural-system boundaries in terms of changes in the landscape structure of the South Florida. The project's value will depend on how well the knowledge encompassed by the modeling system is translated into policy that can be understood and accepted by land users.

The Sustainable South Florida project will probably not evolve into a comprehensive model, but rather will be a collection of alternative scenarios, each of which can be considered an integrated model in itself. The approach we are considering, which can both integrate complex ecological knowledge and represent it in a form that can be presented to policy makers and their constituents, is to develop a computer-based modeling environment in which to run the scenarios (Berry et al. 1996, Coulson et al 1989, Flamm & Turner 1994).

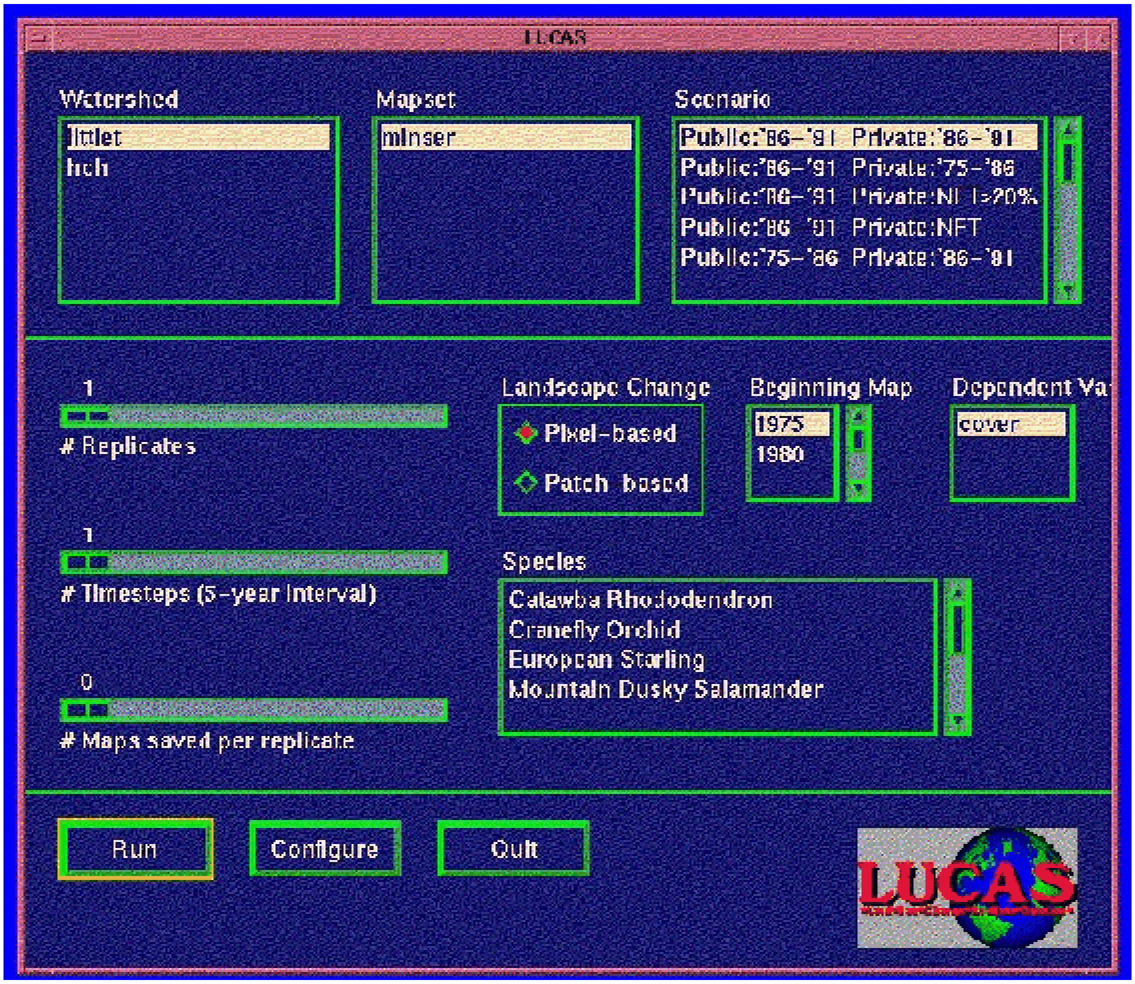


Fig. 4. Set-up menu for running simulation experiments using the Land-Use Change Analysis System (LUCAS).

An example of a computer-based, scenario-driven modeling environment is the Land-Use Change Analysis System or LUCAS (Flamm & Turner 1995, Berry et al. 1996). LUCAS was developed during a U. S. Man and the Biosphere Project in the Computer Science Department at the University of Tennessee. It was designed to run management scenarios so that decision makers could visualize and analyze possible ramifications of today's actions. Figure 4 shows one of the windows that appears on the screen during a simulation exercise. The user selects a scenario to run, the number of timesteps and replicates per timestep, how many maps to save, and which effects to evaluate. Simulations are then run, and the results are displayed in the graphics window, and a variety of spatial indices are calculated.

Why use a system like LUCAS for a large-scale modeling effort like the Governor's Commission on a Sustainable South Florida? First, it's an excellent tool for interdisciplinary work because it can handle many types of information including spatial and tabular data, heuristics, and analytical models. LUCAS is also a warehouse of information, storing existing models and data that can be accessed during the development of future scenarios. Because LUCAS's structure is modular, it is not too difficult to incorporate new models or data into it. And finally, through its user-friendly graphical interface, we can represent simulation results in ways that allow users to explore management alternatives and ultimately to be more effective in formulating and implementing land-use policies.

The examples presented represent spatial models of different complexity and scale, and were designed for different target users. Each can be useful in influencing local land-use decisions, including those made by individuals, in public meetings, in permitting reviews, and in court challenges. However, introducing ecological models into land-use management is especially challenging because of the range of education and values of the people involved. The example of the propeller scarring was a simple map of what's present that was easily integrated into land-use management. The remaining models are less traditional, more complex, and more difficult to integrate into land-use decision making. Challenges to modelers associated with integrating these more complex land-use models into local land-use decisions might include the following: understanding social dynamics among stakeholders, understanding how humans value nature, being able to effectively communicate the concept of uncertainty, acknowledging the role of the courts in environmental management, and educating the public about their impacts on the landscape. In addition, landscape ecological research needs to expand its involvment beyond multidisciplinary land-use modeling into the sociology of how humans use ecological knowledge so that we can begin to investigate how land-use models influence land use.

References

Berry, M.W., Flamm, R.O., Hazen, B.C., and MacIntyre, R.M. (1996) The land-use change and analysis system (LUCAS) for evaluating landscape management decisions. IEEE (In Press).

Coulson, R.N., Saunders, M.C., Loh, D.K., Oliveria, F.L., Drummond, D., Barry, P.J., and Swain, K.M. (1987) Knowledge system environment for integrated pest management in forest landscapes: The southern pine beetle (Coleoptera: Scolytidae). Bull. Entomol. Soc. Am. 34:26-33.

Flamm, R.O. and Turner, M.G. (1994) GIS applications perspective: multidisciplinary modeling and GIS for landscape management. In: V. Alaric Sample [ed.], Remote Sensing and GIS in Ecosystem Management. Island Press, Washington, DC.

Sargent, F.J., Leary, T.J., Crewz, D.W., and Kruer, C.R. (1995) Scarring of Florida's seagrasses: assessment and management options. Florida Marine Research Insitute Tech. Rpt. TR-1.

Richard O. Flamm Associate Research Scientist
Florida Department of Environmental Protection, Florida Marine Research Institute
100 8th Avenue, SE, St. Petersburg, Florida, 33701-5095
Phone: (813)-896-8626, ext 3030
Fax: (813)-823-0166
email: flamm@manatee.fmri.usf.edu

Thomas Maxwell and Robert Costanza

Facilitating High Performance, Collaborative Spatial Modeling

Protecting and preserving our natural life-support systems requires the ability to understand the direct and indirect effects of human activities over long periods of time and over large areas. Computer simulations are now becoming important tools f or investigating these interactions. Achieving a realistic assessment of the impacts of alternative management policies on ecosystem health requires very complex, spatially articulated ecosystem models. The data and hardware barriers to building these m odels have been eroding with the increasing availability of remote sensing data and GIS systems to manipulate it, and the development of parallel/distributed computer systems. Current research is now focusing on alleviating the conceptual barriers to wid e utilization of these tools in a collaborative environment.

In this paper we discuss new modeling technology which supports 1) modular, hierarchical model construction and archiving/linking of simulation modules, 2) graphical, icon-based model construction, 3) transparent distributed computing, and 4) integrating disparate space-time representations. It is hoped that this type of infrastructure will open the simulation arena to a much wider set of participants, and facilitate the application of computer modeling to the study of complex multi-scale processes in sup port of policy making on all levels.

---

1. Introduction

Economic activity, climate, populations, technology, and environmental change are all intimately intertwined. There are many signs that the collective global economic activity is dramatically altering the self-repairing aspects of the global ecosystem . Our ability to change the economic and the ecological systems, and the rate of spread of the impacts of these changes far exceeds our ability to predict the full extent of these impacts. Protecting and preserving our natural life support systems requ ires the ability to understand the direct and indirect effects of human activities over long periods of time and over large areas. Computer simulations are now becoming important tools to investigate these interactions, but our modeling and understanding of these systems has been largely isolated and unconnected in disciplinary specialties.

We believe that it is now possible, because of recent developments in data acquisition and display, information-sharing technologies, and large-scale computing and modeling, to build computer simulation models that cover the salient features of the wor ld's climate, economy, and ecosystems, and the major interactions among them. The combined global ecological economic models will provide both an integrated conceptual framework, and a practical tool allowing researchers from many disciplines to collab orate effectively in order to produce effective answers to the questions mentioned above. To achieve this goal it will be necessary to mobilize this scholarly community within a worldwide "collaboratory" based on new electronic information sharing techn ologies (Smarr 1994), bringing together leaders in advanced computation and software development with leaders in global ecological and economic data collection and modeling. The following sections discuss an on going computational development program in support of this collaborative spatial modeling effort.

2. Supporting Collaborative Modeling

Spatially explicit modeling of ecological-economic systems is essential if one's modeling goals include developing a relatively realistic description of past behavior and predictions of the impacts of alternative management policies on future system be havior (Risser, Karr et al. 1984 ; Costanza, Sklar et al. 1990 ; Sklar and Costanza 1991). There exists a rich set of research problems associated with the impl ementation of computer based collaborative technologies for high performance spatially-articulated ecological economic modeling. Five important areas of ongoing research and development are integrated support for 1) modular, collaborative model developme nt, 2) transparent access to high performance computing resources, 3) graphical display & manipulation of model structure and dynamics, 4) multiple spatial representations, and 5) multiple dynamic modes.

2.1 Collaborative Modular Modeling

Development of ecosystem models in general has been limited by the ability of any single team of researchers to deal with the conceptual complexity of formulating, building, calibrating, and debugging complex models. The need for collaborative model b uilding has been recognized (Goodall 1974 ; Acock and Reynolds 1990) in the environmental sciences. Realistic ecosystem models are becoming much too complex for any single group o f researchers to implement single-handed, requiring collaboration between species specialists, hydrologists, chemists, land managers, economists, ecologists, and others. The current generation of models tend to be "idiosyncratic monoliths that are compre hensible only to the builders" (Acock and Reynolds 1990). Communicating the structure of the model to others can become an insurmountable obstacle to collaboration and acceptance of the model. Policy makers are unlikely to trust a model they don't understand.

A well-recognized method for reducing program complexity involves structuring the model as a set of distinct modules with well-defined interfaces (Gauthier and Ponto 1970 ; Goodall 1974 ; Acock and Reynolds 1990 ; Silvert 1993) Modular, hierarchical model structuring is well developed in the context of discrete event modeling (Zeigler 1976), bu t has received comparatively little development in the realm of continuous modeling (Goodall 1974 ; Cellier 1991 ; Silvert 1993). Ecosystem models with a modular hierarchical structure should be closer to natural ecosystem structure than procedural models (Goodall 1974 ; Silvert 1993), since the component populations of ecosys tems are themselves complex hierarchical systems with their own internal dynamics. Modular design facilitates collaborative model construction, since teams of specialists can work independently on different modules with minimal risk of interference. Mod ules can be archived in distributed libraries and serve as a set of templates to speed future development. The inheritance property of object-oriented languages allows the properties of object-modules to be utilized and modified without editing the archi ved object. A modeling environment that supports modularity could provide a universal modeling language to promote worldwide collaborative model development.

2.2 High Performance Computing

Tremendous computational resources are required to integrate the equations of a large spatial model in a reasonable amount of computer time. Large models typically require supercomputers for efficient execution. This class of models is a near ideal a pplication for parallel processing since a typical model consists of a large number of cells that can be simulated semi-independently. Each processor can be assigned a different subset of cells, and most interprocessor communication is nearest-neighbor only. Despite their great promise and increasing availability, parallel architectures have not found much usage in the life sciences. The major barrier to wide acceptance of these techniques has been the difficulty of programming and debugging large par allel programs, and reluctance on the part of scientists to invest time in learning new languages and architectures. Model builders must usually make a substantial time investment learning a new language or development system when beginning work on a par allel computer.

2.3 Graphical Display

A second step toward reducing model complexity involves the utilization of graphical, icon-based module interfaces, wherein the structure of the module is represented diagramatically, so that new users can recognize the major interactions at a glance. Scientists with little or no programming experience can begin building and running models almost immediately. Inherent constraints make it much easier to generate bug-free models. Built-in tools for display and analysis facilitate understanding, debugg ing, and calibration of the module dynamics.

One major advantage of this graphical approach to modeling is that the process of modeling can become a consensus building tool. The graphical representation of the model can serve as a blackboard for group brainstorming, allowing policy makers, scien tists, and stakeholders to all be involved in the modeling process. New ideas can be tested and scenarios investigated using the model within the context of group discussion as the model grows through a collaborative process of exploration. When applied in this manner the process of creating a model may be more valuable than the finished product.

2.4 Multiple Spatial Representations

Building realistic spatially explicit ecological-economic models requires the integration of multiple spatial data structures in a single model. For example, variables such as elevation and vegetation cover may be require a grid representation, while entities such as roads, rivers, and canals may favor a vector representation. An "area" representation may be most appropriate for lumped-parameter models that may be embedded in a spatial grid, such as a spatially-aggregated lake model that covers mult iple grid cells in a landscape. Other objects may be represented as mobile points, such as entities that can wander around in the landscape. These and other spatial data structures should be implemented in the modeling environment, and the details of li nking, transferring data between, and decomposing (over multiple processors) spatial representations should be transparent to the modelers.

2.5 Multiple Temporal Modes

Building realistic ecological-economic models requires the integration of multiple dynamic modes in a single model. For example, many processes are best represented using differential equations, others are best represented using event-based simulation , and others, such as input-output economic models, use a "black-box" or look-up table implementation. Some processes, such as storm events, are best handled with a hybrid approach. Since continuous (differential-equation based) simulation can be emulat ed in a discrete event framework (but not vice-versa), the underlying temporal dynamics of the simulation environment should be event based, but structured to efficiently emulate continuous systems.

3. Spatial Modeling Environment

In an attempt to address the conceptual and computational complexity barriers to spatio-temporal model development, we have developed a spatial modeling environment (SME), which links icon-based gra phical modeling environments with parallel supercomputers and a generic object database (Costanza and Maxwell 1991 ; Maxwell and Costanza 1994 ; Maxwell and Costanza 1995 ; SME2 1995). This system allows users to create and share modular, reusable model components, and utilize advanced parallel computer architectures without having to invest unnecessary time in computer progr amming or learning new systems. The following sections give a brief description of the current design of the SME. A more detailed description can be found in the web page (SME2 1995).

The SME design has arisen from the need to support collaborative model development among a large, distributed network of scientists involved in creating a global-scale ecological/economic model. It is intended that it's design be progressively more in clusive of the full range of relevant ecological/economic modeling activities. In the interest of maximizing accessibility to a distributed network of collaborators, the system is designed to support a range of platforms, both in the front-end developmen t environment and in the back-end parallel computing environment. We are thus led to the formulation of a three-part Modelbase-View-Driver architecture. The three components are displayed in Fig. 1 and described below.

3.1 View

The View component of the SME is an graphical, icon-based simulation environment used to construct, run, calibrate, and test biological/ecological/ economic modules in a desktop environment. This component is represented by an off-the-shelf applicati on such as STELLA (HPS 1995) or Extend (Extend 1995)

3.2 Modelbase

In the next step toward constructing a spatial model, the Module Constructor translates the View ecosystem component modules into Module objects defined in our text-based Modular Modeling Language (MML). An example of a simple STELLA model simulating a predator-prey system and the corresponding set of MML objects is discussed in the web page (MML 1995). The MML objects can then be archived in the ModelBase to be accessed by other researchers, and/or used immediately to c onstruct a working spatial simulation. Many MML objects can be combined hierarchically in the MML. This MML hierarchy can then be converted by the Code Generator into a C++ object hierarchy within the spatial modeling environment (SME), where it can dri ve a spatial simulation.

The MML is designed to capture only the relevant dynamics of the simulation module being constructed, and leave out all implementation-specific details. For example, the features that can be represented in the MML include dynamics of growth, death, an d transformation of biological/ecological entities, fluxes of water, nutrients, pollutants, etc., and the internal decision and learning processes of biological agents. The features that are not represented in the MML include the spatio-temporal implemen tation of the modules, input and output of model data, and the distribution of the model over a set of processors. These features are configured in the code generation process and implemented in the distributed simulation environment (next sections).

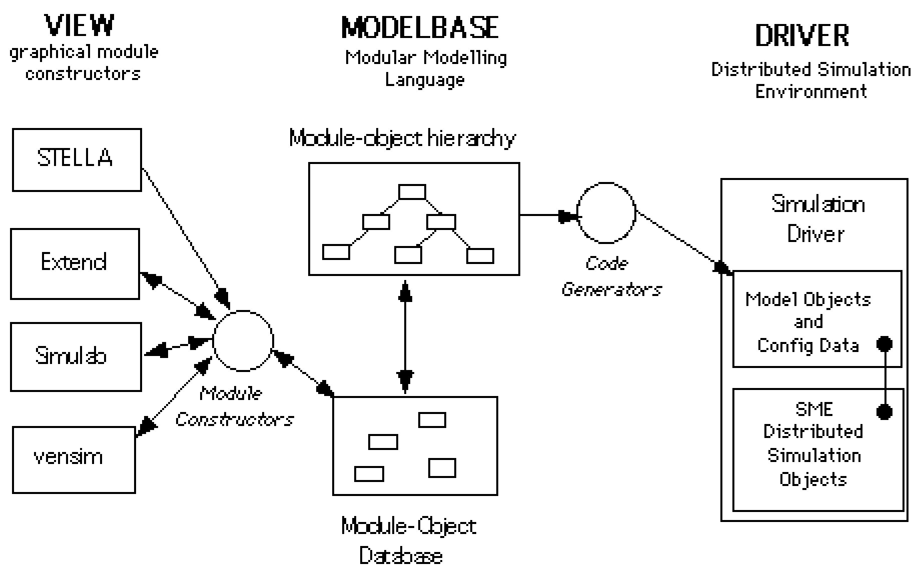


Figure 1: The ModelBase-View-Driver Architecture.

3.3 Code Generators

The Code Generators convert a MML object hierarchy into a C++ object hierarchy which is incorporated into the simulation driver application to create a spatial simulation. The user customizes the set of objects generated by entering information into a set of configuration files that are initially generated by the Code Generator. An example of a simple configuration file is discussed in the web page (MML 1995). In the final version a menu-driven interface will be provide d to facilitate this configuration step.

During the configuration step the user specifies the additional information that is required to transform the MML object into a dynamic simulation object. The information entered falls into several general categories, including:

- Space-time implementation: In this step, each MML object is associated with a frame, which specifies its space-time implementation. A frame is a C++ object which specifies the topology of the spatial implementation of the module, methods for intera cting with and transferring data to other frames, and temporal methods for handling the passing of time. The driver geometry object (MML 1995) maintains a catalog of available frames. Examples of available frames include t wo-dimensional grids (e.g., for landscapes), graphs and networks (e.g., for river, canal, or neural networks), and agents (e.g., for individual agents moving about in the landscape). The user specifies a frame type as well as a (set of) GIS map(s) that t he frame will read at runtime to configure itself.
- Input/Output configuration: In this step the user specifies input data types (e.g. map, timeseries, database parameter, etc.) for the simulation's input variables. Input configuration must be done in code generation phase because the Code Generator uses this information (together with the variable dependency graph) to derive information (e.g. spatio-temporal variability, etc.) about the simulation variables. Output configuration is done at runtime, although default values can be specified in the C G configuration files.

3.4 PointGrid Library

The PointGrid library (PGL) is a set of C++ distributed objects designed to support computation on irregular, distributed networks and grids. It contains the core set of objects on which the SME Driver is constructed. The PGL builds spatial represen tations from sets of Point objects (see below) with links. It transparently handles: 1) creation and decomposition (over processors) of Point Sets, 2) mapping of data over and between Point Sets, 3) Iteration over Point Sets and Point Sub-Sets, 4) data access and update at each Point, and 5) swapping of variable-sized PointSet boundary (ghost) regions. Some of the important PGL classes are:

- Point: Corresponds to a cell in a GIS layer.
- Aggregated Point: Corresponds to a cell in a coarser resolution GIS layer.
- PointSet: A set of Points with links (grid or network).
- DistributedPointSet: A PointSet distributed over processors with variable-sized boundary (ghost) layers.
- Coverage: Mapping from a DistributedPointSet to the set of floats.

For example, consider the study area for the Patuxent Landscape Model, displayed in Figure 2. Each cell in the (non-black) study area is represented by a Point Object. Each cell (Point) in the blue-green/gray are a is also part of the PLM river network. The PLM utilizes two PointSets: 1) A base grid PointSet which includes all Points in the study area with links to nearest neighbors in eight directions, and 2) a Network PointSet which includes all Points in the r iver network, each Point having a single link to it's downstream cell.

The white lines on the figure show how a DistributedPointSet object distributes the Points among four processors. The distribution algorithm attempts to allocate equal sub-sets of the study area (base grid) to each processor; all other PointSets inherit t he same decomposition.

A Coverage object associates a floating-point number with each Point in it's PointSet. Each

PointSet contains methods for efficiently 1) iterating over Points in the Set, 2) accessing neighboring Points in the Set, 3) accessing and updating associated Cov erage values, 4) and mapping Coverages between PointSets.

3.5 Driver

The driver is a distributed object-oriented simulation environment which incorporates the set of code modules that actually perform the spatial simulation on the targeted platform. It is implemented as a set of distributed C++ objects linked by messag e passing, layered on top of the PointGrid library. Some of the important driver classes are:

- Module class: The CodeGenerator Application converts each module in the MML model description into a Module Object in the Driver. Each Module Object has a set of Variable Objects and a Frame Object. It also has a set of methods for responding to s imulation events such as initialize and update.
- Frame class: A Frame Object is a driver object which specifies the topology of the spatial implementation of a Module Object, including methods for interacting with and transferring data to other frames. A frame has a list of Point Objects (POs), wi th each PO corresponding to a cell in the frame's map region, which includes a partition of the study area handled by the current processor plus a communication buffer zone. The driver maintains a catalog of available frames, which includes two-dimension al grids (e.g., for landscapes), graphs and networks (e.g., for river, canal, or neural networks), areas (e.g. for embedded lumped-parameter models) and point collections (e.g., for individual agents moving about in the landscape).
- Variable class: The CodeGenerator Application converts each variable declared in the MML model description into a Variable Object in the Driver. The Variable class is a specialization of the Coverage class, which encapsulates a mapping from the set of Point Objects owned by the Module's Frame Object into the set of floating point numbers.
- Simulation services: The driver also incorporates a set of objects for handling various simulation services, such as input and output of data from databases and GIS, network services for distributed computing, and display/visualization services for viewing simulation output.

4. A Modular Modeling Language

The core of the SME is our text-based Modular Modeling Language (MML 1995). The MML is designed to capture only the relevant dynamics of the simulation module being constructed, and leave out all implementation-specific d etails. For example, the features that can be represented in the MML include dynamics of growth, death, and transformation of biological/ecological entities, fluxes of water, nutrients, pollutants, etc., and the internal decision and learning processes o f biological agents. The features that are not represented in the MML include the spatio-temporal implementation of the model, input and output of model data, and the distribution of the model over a set of processors. These features are implemented by the Code Generator and the simulation drivers.

As an example of MML Module development, the table (MML 1995) displays a set of MML modules corresponding to the STELLA model in the figure, which simulates a simple predator-prey system. These equations were generated by the ModuleConstructor application from the STELLA model's equation export file. There are three modules displayed, `Hare_Module` (encapsulating hare dynamics), `Lynx_Module` (encapsulating lynx dynamics), and `PredPrey_Module` (representing the linked model). Each Module has a set of Variable Objects, which can be internal (declared with the Variable command) or input from another Module (declared with the Input command). All internal Variables can serve as exports to other modules. The higher level Module `PredPrey` incorporates the `Connection` co mmands which link outputs of one module to inputs of another.

Each Module also has a set of `Event` commands, which specify the actions that the module executes in response to various simulation events. Shown in the example are `Initialization` (performed at the beginning of a simulation) a nd `Update` (performed at each timestep of the simulation) events. The module's responses to these events generate the dynamics of the simulation.

The table (MML 1995) displays the configuration file associated with this model. A default version of this file is generated by the ModuleConstr uctor application. The lines beginning with '$' configure Module objects, and lines beginning with '*' configure Variable objects. The `g(2D,file)` command links the Hare and Lynx Modules to 2D grid frames which are initialized with map file. The `d(file)` commands declare the associated variables to be map input objects, which are initialized with map file. The `A()` commands configure animation output.

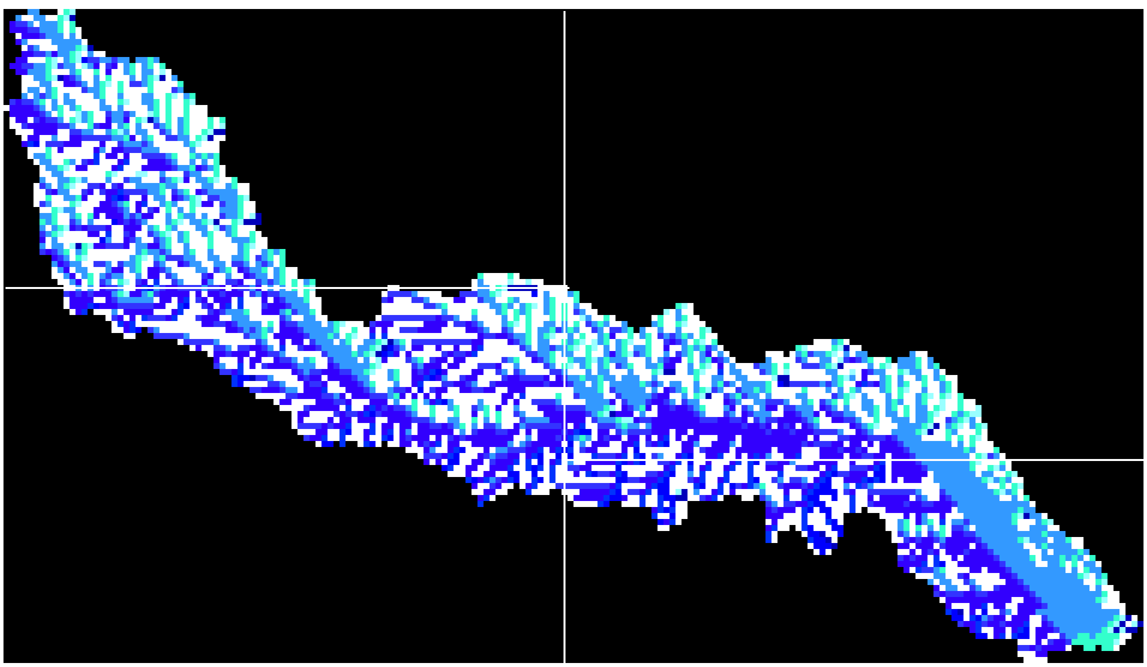


Figure 2: The Study Area for the Patuxent Landscale Model.

5. Patuxent Landscape Model

The current applications of this framework include the Everglades Landscape Model (ELM 1995), spatial modeling at the UIUC GMSLab (GMSLab 1995) , and the Patuxent Landscape Model (PLM 1995), a regional landscape simulation model that can address the effects of different management and climate scenarios on the ecosystems in the Patuxent Watershed. The PLM is being de veloped as part of the Ecological Ecosystem Models for Evaluating the Interactive Dynamics of the Patuxent River Watershed and Estuary Project funded by the Chesapeake Bay Research & Monitoring Division, Maryland Department of Natural Resources. The PLM c ontains about 6,000 spatial cells each containing a dynamic simulation model (based on the GEM model (Fitz, DeBellevue et al. 1995)) containing approximately 20 state variables partitioned into 14 modules. It uses two frames, a 2D grid frame (for modules such as Algae, Consumers, Nitrogen, Hydrology, Macrophytes, Detritus, etc.) and a tree-network frame for the River modules. The model is being calibrated with data from 1973 and 1985, and run for a scenario analysis period from 1985 to 2020 with selectively variable time steps from hourly to daily depending on forcing function dynamics. Application of this model in the Patuxent watershed is expected to allow extensive analysis of past and future management options, and will form the basis for future application to other areas in the Chesapeake Bay watershed.

6. Conclusions

We believe that effectively managing human affairs through the next century will require extremely complex and reliable computer models. Widespread utilization of modeling environments supporting graphical, hierarchical/modular design linked to advanc ed computing resources may be essential in facilitating reliable, economical model construction. General adoption of this paradigm will support the development of libraries of modules representing reusable model components that are globally available to model builders, as well as making advanced computing resources available to users with little computer expertise.

7. References

Acock, B. and J. F. Reynolds (1990). Model Structure and Data Base Development. *Process Modeling of Forest Growth Responses to Environmental Stress.*. R. K. Dixon, R. S. Meldahl, G. A. Ruark and W. G. Warren. Portland, OR, Timber Press.

Cellier, F. E. (1991). *Continuous System Modeling.*. New York, NY, Springer-Verlag.

Costanza, R. and T. Maxwell (1991). "Spatial Ecosystem Modeling Using Parallel Processors." *Ecological Modelling* **58**: 159-183.

Costanza, R., F. H. Sklar, et al. (1990). "Modeling Coastal Landscape Dynamics." *BioScience* **40**: 91-107.

ELM (1995). Everglades Landscape Model. URL: http://kabir.umd.edu /Glades/ELM.html.

Extend (1995). Extend Simulation Software. San Jose, CA, Imagine That Inc.

Fitz, H. C., E. DeBellevue, et al. (1995). "Development of a General Ecosystem Model (GEM) for a range of scales and ecosystems." *Ecological Modelling* (in press).

Gauthier, R. L. and S. D. Ponto (1970). *Designing Systems Programs*. Englewood Cliffs, NJ, Prentice-Hall.

GMSLab (1995). Spatial Modeling at the Uniiversity of Illinois GMSLab. URL: http://ice.gis.uiuc.edu/index.html.

Goodall, D. W. (1974). The Hierarchical Approach to Model Building. Proceeding of the First International Congress of Ecology, Wageningen, Center for Agricultural Publishing and Documentation.

HPS (1995). STELLA: High Performance Systems. URL: http://www.hps-inc.com/.

Maxwell, T. and R. Costanza (1994). Spatial Ecosystem Modeling in a Distributed Computational Environment. *Toward Sustainable Development: Concepts, Methods, and Policy*. J. van den Bergh and J. van der Straaten. Wash ington, D.C., Island Press: pp. 111-138.

Maxwell, T. and R. Costanza (1995). "Distributed Modular Spatial Ecosystem Modelling." *International Journal of Computer Simulation: Special Issue on Advanced Simulation Methodologies* **5**(3): 247-26 2.

MML (1995). Modular Modeling Language Example. URL: http://kabir.umd.edu/SMP/MVD/T2.html.

PLM (1995). Integrated Ecological Economic Modeling. URL: http://kabir.umd.edu/PLM/PLM_Proj.html.

Risser, P. G., J. R. Karr, et al. (1984). Landscape Ecology: Directions and Approaches, Illinois Natural History Survey, Champaign, IL.

Silvert, W. (1993). "Object-Oriented Ecosystem Modeling." *Ecological Modeling* **68**: 91-118.

Sklar, F. H. and R. Costanza (1991). The Development of Dynamic Spatial Models for Landscape Ecology. *Quantitative Methods in Landscape Ecology*. M. G. Turner and R. Gardner. New York, NY, Springer-Verlag. **82**: 239-288.

Smarr, L. (1994). Personal Communication.

SME2 (1995). Spatial Modeling Environment, version 2 alpha. URL: http://kabir.umd.edu/SMP/MVD/SME2.html.

Zeigler, B. P. (1976). *Theory of Modeling and Simulation.*. New York, N.Y., Wiley.

Author Information

Tom Maxwell

University of Maryland,
Institute for Ecological Economics
Box 38, Solomons, Md. 20688
maxwell@kabir.cbl.cees.edu
http://kabir.cbl.cees.edu/Tom/Maxwell.html

Robert Costanza
University of Maryland,
Institute for Ecological Economics
Box 38, Solomons, Md. 20688
costza@cbl.cees.edu

Cees G. Wesseling, Willem P.A. van Deursen, Peter A. Burrough

A spatial modelling language that unifies dynamic environmental models and GIS

This paper deals with approaches for linking GIS and models with a focus on environmental applications. Although examples of the integration of models in current GIS programs exist, optimal integration is in our view reached only if the GIS enables the construction of any dynamic model using the tools included in the GIS. This aim can only be accomplished if the design and implementation of the analytical engine and the spatial database are properly integrated.
The approach taken was to design a GIS-based modelling language that allows the dynamic behaviour of environmental systems to be modelled without the burden of technical implementation and algorithm details.
Equally important was to keep the language understandable and easy to learn. This design goal was met by creating a minimal set of standard functions which are the building blocks for a large number of environmental models.

Introduction

The contributions to the Breckenridge 1993 conference showed numerous successful examples of coupling GIS and specific environmental models. Several papers discussed the issue of generic coupling GIS and models from different points of view. Maidment (1993) approached the issue from a GIS perspective with questions like: "Is it possible to use or adapt existing GIS functions to carry out environmental modelling within a GIS system?". Leavesley (1993) puts the model in the central role by implementing the Modular Modular Modeling System (MMS) that holds model components for interactive model construction, with GIS as an interface for spatial model parameters. Kemp (1993) investigated the underlying spatial data models used in GIS and environmental modelling and proposed a data type mechanism of data models for a sound conceptual coupling of both worlds. Common themes in these papers are the recognition of the need for generic components for model building and the construction of a framework that includes the temporal dimension of dynamic environmental models. A prerequisite for the use of generic components is the implementation and identification of data types to verify whether components can be coupled or not.

This paper discusses the development of a Dynamic Modelling language to construct dynamic models using a GIS database that provides strongly typed temporal spatial data. The language and database are part of the new version of PCRaster (Van Deursen 1995). An elaborate overview of the language is given in Wesseling *et al.* (1996). In this paper we emphasize the underlying motivations and design issues of PCRaster's major components: the language, the analytic engine and associated database.

The Database

Computer languages are constructed around data types and data models. In the case of a special purpose programming language for building dynamic models in a GIS environment, the data manipulated by the language are stored in the GIS database, so the language design and the GIS data model design must be consistent with each other.
The question *how* to store data seems inevitable when discussing GIS and data models, as can be concluded from the everlasting raster versus vector discussion. More important however, is the question *what* is stored in the database, and what the characteristics of the entities stored are.
For example, a model may include soil type maps and land use maps. A cross tabulation of these entities is possible whether they are stored as rasters or polygons, but a multiplication of soil class numbers with land use numbers cannot produce anything useful since both entities are nominal data types. Even a comparison (equality) would be nonsense since both nominal data types represent different entities. To overcome these difficulties, and for a functional support of the operations, PCRaster includes type information in the database, and a strong type checking mechanism is applied to the GIS operations.

The PCRaster database can hold raster maps, point data, tables for cross tabulation between maps and time series for representation of attributes changing over time. All these data representations have types attached. These types describe the special properties of the entities modelled. The following data types are recognized:

- boolean
- nominal, subtyped by its legend
- ordinal, subtyped by its legend
- scalar field
- directional field
- local drain direction (ldd)

Scalar fields have a continuous linear scale and are used to describe intensities and potentials of physical fields, such as temperature, population density and precipitation. Directional fields apply to attributes that have a circular continuous scale, such as aspect of the terrain.
The local drain direction (ldd) data type has been introduced to provide for the definition of the direction of potential flow. It is a data type that supplies the raster database with the topological linkages needed by operators that describe lateral fluxes of fluids or materials. For a further discussion on the construction of this data type see Jenson & Domingue (1988) and Moore *et al.* (1993).
The dynamic behaviour of the database, needed for dynamic models, is modelled by time series indexed on time and location, and by stacks of map layers representing the status of the model at different time steps.

Operations on the database entities can only be carried out if the data have the correct data type for the operator concerned. Such a strong type checking scheme assists error detection and helps the user when conceptualizing his ideas about the entities used. An additional advantage of typing the entities is that more intelligence, such as polymorphic behaviour, can be incorporated in the operators. For example, an interpolation operator can automatically

choose the nearest neighbour algorithm for ordinal data and a bilinear interpolation for scalar data.
In the next section we present the language that embeds all analytical operators used for Dynamic Modelling.

The Building Blocks

The PCRaster spatial modelling language is an extension of the ideas behind Map Algebra and the Cartographic Modelling Language proposed by Berry and Tomlin (Berry 1987, Tomlin 1990) but includes new ideas of iterations used in Dynamic Modelling. It follows the same approach as Map Algebra in the sense that it provides a set of generic operators, which can be used as primitives for the models. But the notation form is different. Tomlin proposes a natural language that is understandable for a large group of users with no former experience in computer programming. In our case the users are mainly the developers of dynamic models, who are more comfortable with compact mathematical notations. Therefore, the syntax of the language is based on mathematical equations where each equation assigns the value of an expression to a single output. The equation for sediment transport capacity derived by Kirkby (1976)

G = C * Q^d * sin(S) * 10^-3

is applied to the raster maps `CoverFactor` (C), `FlowVolume` (Q) and `Slope` (S) with an exponent (d) of 1.7 using the command

```
TransportCapacity = CoverFactor*FlowVolume**1.7 *sin(Slope)*10E-3;
```

The command above is a point operation, where a new value for each location, a grid cell, is derived from different attribute values on that same location only. Global operators, where a new value for each location is derived from different attribute values on (possible) different locations, are also modelled as mathematical functions.
The set of available global operators in PCRaster is very extensive in comparison to the range of operations generally considered as Map Algebra. A rich suite of geomorphological and hydrological operators is available. These include functions for hillslope and catchment analysis, and the definition of topology for modelling transport (drainage) of material over a local drain direction map with routing functions.
Although the language contains over fifty global functions, the set of functions and operators is kept as small as possible. In principle, every possible function that can be constructed from the set of other functions is not included. This point is illustrated by an example where a routing function (`accuflux`) is used to calculate the upstream area of one or more catchments.

The operator `accuflux` calculates the accumulated amount of material that flows out of the cell into its neighbouring downstream cell. This accumulated amount is the material in the cell itself plus the amount of material in all upstream cells of the cell. The topological linkages defining upstream and downstream neighbours is given by a local drain direction map as the first argument. The second argument of `accuflux` is in general a map with material to be transported such as surface water as in the next example:

```
WaterFlow = accuflux( Ldd, WaterAmount ) ;
```

To calculate the upstream area of each location within one or more catchments, the `accuflux` function can be applied as follows:

```
UpstreamArea = accuflux( Ldd, cellarea() ) ;
```

Instead of transporting material, the area of each individual grid cell is transported, resulting in a summation of all cells that make up the different (sub) catchments. Note that `cellarea()` results in a single value. Most arguments of PCRaster operators can be either spatial varying, by supplying a map or an expression that yields a map, or constant at each location denoted by a single value or numeric expression. This feature enhances the generic use of the operators.

The examples described above can be invoked from the command line or can be typed in a static Cartographic Modelling script, which is executed by the PCRaster program CALC. The following section shows how PCRaster provides the functionality to embed these simple statements in a dynamic model that iterates the computations for a series of time steps.

Building Dynamic Models

Dynamic models are constructed by writing scripts containing series of statements. The language has no explicit structures for iteration, although dynamic models do iterate in time. Instead, there are different sections that are controlled by the definition of a timer.
The timer section regulates the duration and time slice of the model through three parameters, starttime, endtime and timeslice. The initial section sets the initial conditions for the model, including maps and non-spatial attributes. These values may be defined with one or more PCRaster operations. The dynamic section defines the operations for each time step that result in a map of values for that time step. Each time step consists of one or more PCRaster operations which are performed sequentially.
The next example shows a simplified model script that incorporates precipitation, infiltration and overland flow.

```
# <- this symbol is typed at the start of a comment line
timer
 1 28 1; # 28 timesteps of 6 hours

initial
 # coverage of meteorological stations for the whole area
 RainZones = spreadzone(RainStations,0,1);
 # create a map of infiltration capacity (mm/6hours),
 # based on a soilmap
 InfiltrationCapacity = lookupscalar(SoilInfiltrationTable,SoilType);

dynamic
 # add rainfall to surface water (mm/6hours)
 SurfaceWater = timeinputscalar(RainTimeSeries,RainZones);
 # compute both the runoff and actual infiltration
 Runoff, Infiltration =
  accuthresholdflux,
accuthresholdstate(Ldd,SurfaceWater,InfiltrationCapacity);
 # output runoff at each timestep for selected locations
 report SampleTimeSeries = timeoutput(SamplePlaces, Runoff);
```

In this example, rainfall is the dynamic input. `RainTimeSeries` is a time table with the precipitation measured at several meteorological stations. `RainZones` does not represent the location of the stations but the area of pixels for which the measurement at that station is the best estimation of the actual precipitation at each pixel. Such a map can easily be computed from a station location map, as done in the `initial` section with the application of `spreadzone`. The `RainZones` map denotes for each pixel the column number in the time table.

At each time step, `timeinputscalar` reads a row associated with the current time step and returns a map containing the column values as defined by `RainZones`. Note that the current time step is an implicit argument to all dynamic functions, such as `timeinputscalar` and `timeoutput`.

The second operation in the `dynamic` section transports the `SurfaceWater` over the local drain direction map `Ldd`. This is done with the `accuthreshold` operator which is one of the transport functions (named accu-operators) that accommodate transports restricted by certain transport functions. The `accuthreshold` operator transports water only once the infiltration capacity, given on the map `InfiltrationCapacity`, is exceeded. It results in two maps: a map with the actual infiltration (`Infiltration`) and a map with the amount of overland flow (`Runoff`). The map `InfiltrationCapacity` is calculated in the `initial` section on basis of a soil map (`SoilType`) and a cross table (`SoilInfiltrationTable`) that gives for each soil type the infiltration capacity.

The last statement of the example creates time series from certain locations in the `Runoff` map. These locations are identified by the `SamplePlaces` map. Each non zero value in the `SamplePlaces` map stands for a column in the time series (`SampleTimeSeries`).

A PCRaster script that contains a dynamic section does not write any results to the database unless the keyword `report` is added. This prevents the surplus storage of intermediate results. Thus, typing the `report` before a statement like `Runoff = accuthresholdflux(...)` creates a stack of raster maps for each time step.

This concludes our brief description of the PCRaster Dynamic Modelling language (See Wesseling *et al.* (1996) for a more detailed description). In the rest of this paper we evaluate our experiences with the PCRaster system and reflect on some advantages, disadvantages and future directions of the system.

Usability of the System

Over a period of 8 years, the PCRaster system has evolved from a tool for teaching basic Map Algebra to a system that is capable of complex Dynamic Modelling in a GIS environment. During this period early prototypes of the language have been used for numerous environmental models. Published examples are LISEM, a physically-based hydrological and soil erosion model on catchment scale (De Roo *et al.*, 1996), RHINEFLOW, a water balance model for the river Rhine (Van Deursen & Kwadijk, 1993) and Calluna, an ecological model for heathland dynamics (Van Deursen & Heil, 1993). RHINEFLOW and Calluna are now both written in the final version of the language, both with less than fifty lines of code.

The emphasis during development of the language was mainly on applications in the field of hydrological surface routing. This has resulted in a language which comprises a rich suite of global operators involving drainage networks.

Since the Dynamic Modelling language elaborates on the concepts of Map Algebra, the learning curve for new users is similar. In PCRaster, the user has one interface to all

analytical capabilities of the package because all user modes are provided by one program. This program (named CALC) uses operators which have exactly the same meaning whether they are applied from the command line, in (static) Cartographic Modelling or Dynamic Modelling scripts. Such consistency aids the user to grasp the underlying concepts step by step.
Once the user is familiar with the system, models can be built quickly. In most cases, the 'source code' of the model resides at the comprehensive abstraction level of one or two lines of source code per process (e.g. interception, infiltration or sediment routing). Different sub-models for individual processes or model parameters can be replaced or changed quickly in the model script so that different scenarios can be computed and compared. The results of these scenarios could be the start of hypothesis generation or testing. All input and output of a model, (stacks of) maps and time series, can be visualized instantly without any further data exchange. Model parameters that are both temporal and spatial are visualized by animation through time.

Even though it is possible to build a wide range of very complex models with PCRaster, it would be arrogant to state that every imaginable model can be built. For example, operators for solving ordinary and partial differential equations, that are used extensively in environmental modelling (Maidment 1993), are not yet directly supported. Also the use of vector fields and associated processes, such as diffusion and advection, are not yet available for general use.
Current experiments in these areas are however hopeful. The PCRaster based model LISEM already uses operators for the kinematic wave model for surface routing (Chow *et al.* 1988) and Newton's iteration method for solving differential equations. New operators and concepts will only be added to the system if they have proven to be generic, relatively easy to use in a particular problem domain and if they can be implemented in an efficient manner.

Furthermore, the user is offered possibilities to add lacking functionality using different coupling approaches (Burrough 1996), termed loose coupling, tight coupling and embedded coupling.
Loose and tight coupling is supported by use of a simple and consistent ASCII based format for point data, tables and time series and an ANSI C library that can create, read and write binary raster maps using a few lines of C code. Additionally raster maps can be exported in ASCII format with every possible layout required for formatted input and output in languages such as Fortran. All modules of PCRaster are command line programs that can be combined with other software in UNIX shell scripts or MS-DOS batch files.
Future plans will make embedded coupling possible. Users can link their own functions into the CALC program with dynamic linked libraries (a.k.a. as shared libraries or DLL's). User defined functions are then available in the Dynamic Modelling language in exactly the same strong typed way as standard functions and operators. LISEM already uses a prototype of this coupling mechanism, to incorporate a modified version of the SWATRE soil water model, which simulates the vertical movement of water in the soil (Belmans *et al.* 1983).

Implementation Issues

During development of the software, we have paid attention to execution speed which, after user friendliness, is the most important requirement for a usable modelling system.

In respect to performance it is important to note that the language is not just another macro language. A PCRaster Dynamic Modelling script is not a list of actions that is sent to separate autonomic modules of the GIS. Instead, a single program (CALC) reads a script entirely, checks it for both syntax errors and data type conflicts and then executes it. This approach yields better error detection facilities and faster execution of the model than macro languages.

Execution of an entire model by one program offered us the opportunity to implement several optimization techniques. Results of calculations are kept in memory for direct use in other calculations. Memory requirements are minimized by live-variable analysis (Aho *et al.* 1986). All operations are programmed in tight loops which make them perfect candidates for loop unrolling. If specialized hardware, such as multiprocessor and vector processing, is available other optimization techniques can be applied as well. All models programmed in PCRaster will gain from such hardware without any modification. It is important to notice that, although optimization techniques require some additional computing, the performance benefits of the techniques are large in spatial modelling. PCRaster optimizes scripts at a symbolic level, only dealing with map names and numerical constants, so computing the optimal execution path is independent of the data size (e.g. number of grid cells, number of time steps), but the benefits do increase with the data size.
Although programming one particular model in Fortran or C will always yield faster execution then using a generic tool kit like PCRaster, we are confident that the performance gap is acceptable. A test with the RHINEFLOW model supports this assertion. The model implemented in C was only 8 times as fast as the model written in the PCRaster Dynamic Modelling language. In general the performance gap decreases if the model makes extensive use of trigonometric (e.g. sine) and logarithmic (e.g. log, power, square root) functions.

In retrospective, the choice of implementing a high level dynamic modelling language has proven effective on both the user interface and the computing interface. The language offers a notation that is very close to the conceptual framework of most users, without the burden of technical implementation and algorithm details. Even on ordinary personal computers, the computing interface is effective in terms of performance, while offering transparent use of specialized hardware when available.

The most frequent asked question when demonstrating the Dynamic Modelling language to other people is: "When will you add a Graphical User Interface?" (often accompanied with names of apparent popular operating systems). The answer is that we are still working on the novel of our story and that the motion picture will add only a few new insights to the story. Apart from that, the motion picture should be called "The NeverEnding Story", which has already been filmed. The author Michael Ende (Die Unendliche Geschichte) decided that he was unhappy with the film's version of his novel, and refused to have his name placed in the opening credits.
In the next section we discuss both the story and the upcoming motion picture.

Future Directions

PCRaster is operational on both UNIX and MS-DOS (80486 or better) platforms with the graphical modules working under X11 and SuperVGA respectively. There are no differences between both versions of the system. In addition to standard GIS functionality, such as data import/export, hardcopy output and graphical display, the PCRaster package includes two extra packages which are linked following the tight coupling approach. These are ADAM (Heuvelink 1993, Wesseling and Heuvelink 1993) for estimation of error propagation in GIS operations and GSTAT (Pebesma 1996ab) for geostatistical modelling, interpolation (kriging), conditional simulation and random field generation.

Although the PCRaster system, in its current form, is already a practical tool to create and maintain many environmental models as well as other spatial models, a number of improvements can be made. Improvements considered are both functional improvements, to support more modelling techniques, and interface improvements to make the system more user friendly.

The most important functional improvement is the implementation of the vector field concept (Kemp 1993). Further research and application of PCRaster in areas like groundwater modelling, where vector based processes such as diffusion and advection are used extensively, is necessary. Generic support for this type of models also requires numerical procedures for solving user defined differential equations.
To execute such numerical procedures efficiently, some other optimization techniques are required. One technique considered, is to translate a Dynamic Modelling script into an ANSI C program that can embed these procedures in the most efficient way. Such a translation is completely hidden from the user. ADAM already uses this translation technique for Monte Carlo simulations applied on Cartographic Modelling scripts with correlated random parameters.

Error propagation analysis is currently possible by combining different tools from the PCRaster system. The Dynamic Modelling language (CALC), has functions for creating single numbers and fields with an uniform or normal distribution. ADAM supports estimation of error propagation in Cartographic Models. GSTAT can be used to generate conditional or unconditional field realizations of random parameters. Output from these tools can be used as input in a Dynamic Modelling script, following the tight coupling mechanism.
A more structured approach would be to add the distinction of deterministic and random parameters in the PCRaster data type system and integrate (un)conditional simulation and error propagation in CALC, the analytical engine of PCRaster. All data types of PCRaster could be subdivided in deterministic and random representations. To name a few possibilities, scalar types (e.g. infiltration capacity) can be represented by a continuous distribution such as normal or log normal. Nominal entities (e.g. land use) can have a probability of belonging to the different classes discerned in a particular nominal type.

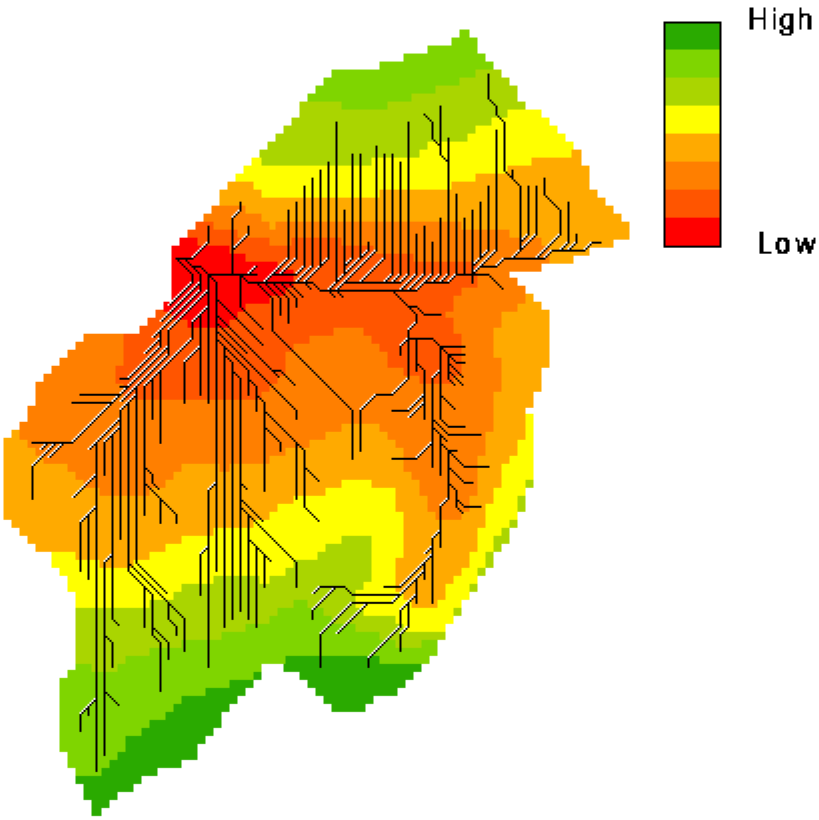
An important interface improvement is to extend the modelling language with a component (or object) based syntax (Van Deursen 1995). Many environmental models are best described as storages that are capable of storing material (e.g. water on surface, water in the soil) and transports of material between such storages (e.g. rainfall on surface, infiltration from surface to soil). A better structure for these models is to define components, link them together and describe the characteristics of each individual component that make up a dynamic system separately. In the future, the Dynamic Modelling language offers an component based syntax to define the components, storages and their connections, called transports. The modelling language that results is based on the system's dynamics approach of Forrester (1968) and is a spatial extension of an approach similar to the STELLA (TM) modelling environment. Note that the component-based description of the model does not define the sequence of execution. The execution order is determined by the interdependencies of the storages and transports. Such an implicit execution order is not appreciated by all users. Other user interface concepts, than an apparently sequential script, must be developed to assist the user in deploying the component based description. A graphical user interface representing a flowchart of storages and transports could be a great improvement.

In general however, we believe that a sound and intuitive language offers a more structured approach than a Graphical User Interface (GUI). Therefore, a GUI should not replace the language interface but instead assist the language interface. This can be done by integrating an editor, manual, examples and a GIS database browser in a single GUI with hyperlink and cut and paste capabilities. The GIS database browser can be used as a way to link model parameters, input and output with maps in the GIS. By making the GIS database aware of models and different scenarios of one model, useful links can be made in the GIS database browser. This enables the user to explore all kinds of spatial and temporal relations and differences between scenarios graphically.

The example model

This section illustrates the example model described in this article with a case study. The study area used is a 600 x 800 m. catchment that is gridded with a cell size of 10 x 10 m. The major drainage direction in the study area is to the NW.

Below: Digital elevation model of the study area with the major flow directions
The flow directions shown here are those with an upstream area of more than 1000 square metres.

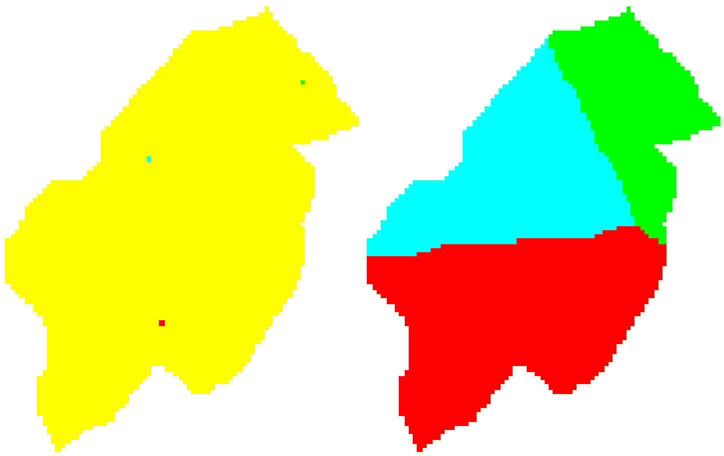


<<Return to article.

spreadzone

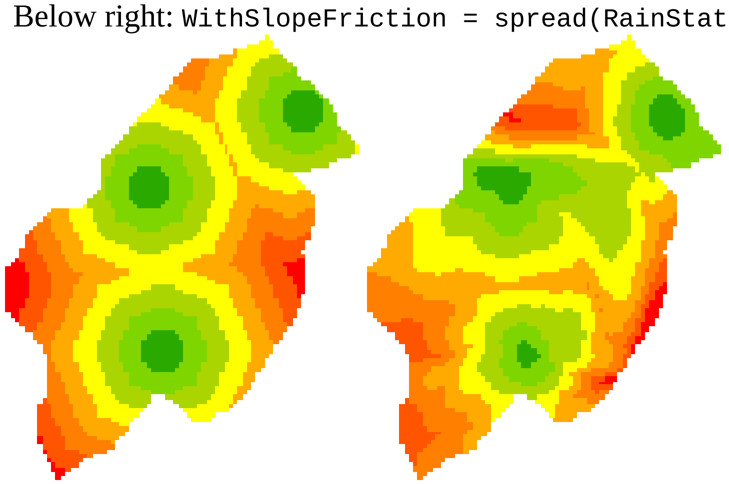
The spread and spreadzone functions of PCRaster both implement the same cost distance mapping algorithm with three arguments: source, initial cost at source and frictional cost. In the model described in the article, these arguments are respectively the map `RainStations` (containing the rainstations, uniquely numbered), the value 0 (identifying no initial cost) and the value 1 (identifying a constant frictional cost of 1 per metre). The function spread returns a map with the travel distance to the nearest (in terms of travel distance) source cell (in this case one of the rainstations). The function spreadzone gives a map with the unique code of the nearest source cell.
The model uses only the function spreadzone. The input map with the rainstations and the resulting map with the rainzones related to the nearest rainstations are shown below.

```
RainZones = spreadzone(RainStations,0,1);
```
Below left: `RainStations`
Below right: `RainZones`



The maps below give the result of the spread function. The left map contains the result of the spread function applied to the `RainStations` map with a frictional cost of 1. It returns for each cell the absolute travel distance (metres) to the nearest rain station. The right map gives the result of the spread function when a spatially varying frictional cost is defined: in this case the slope of the terrain. The result (below right) contains relative travel distances (incorporating the slope of the terrain) from the rain stations. On this relative distance map, cells that can only be reached from a rain station by travelling through steep terrain have a large relative distance value. Cells that can be reached from a rain station by travelling through flat terrain (for instance the valleys) have low values.

Below left: `DistFromStation = spread(RainStations,0,1)`
Below right: `WithSlopeFriction = spread(RainStations,0,slope(Dem))`



<<Return to article.

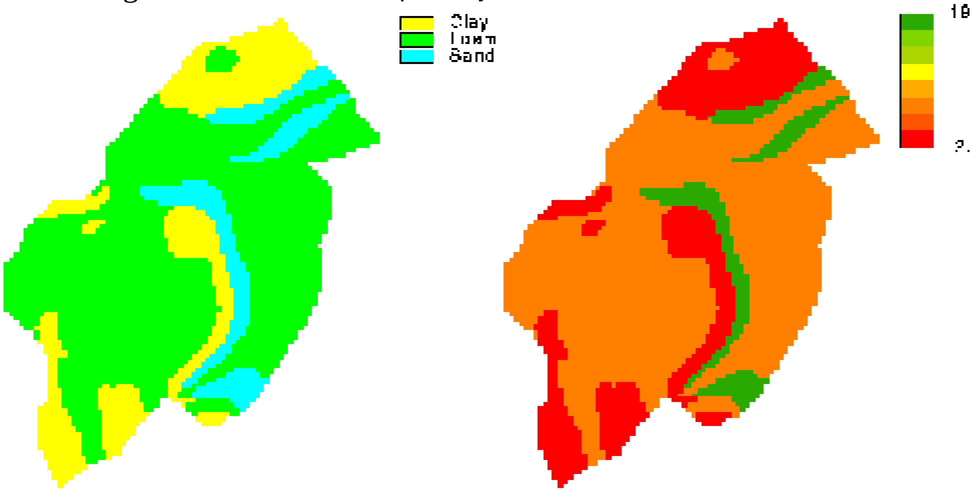
---

lookupscalar

Lookup functions are used to relate tabular data to maps. The first argument is a lookup table that holds a selection criteria in each column and the resulting value in the last column. The example below is used in the model where clay has a map value 1 in the soil map and loam and sand a value 2 and 3 respectively.

```
InfiltrationCapacity = lookupscalar(SoilInfiltrationTable,SoilType);
```
contents of `SoilInfiltrationTable` table:
```
1 2.1
2 8.3
3 19.0
```
Below left: `SoilType`

Below right: `InfiltrationCapacity`



<<Return to article.

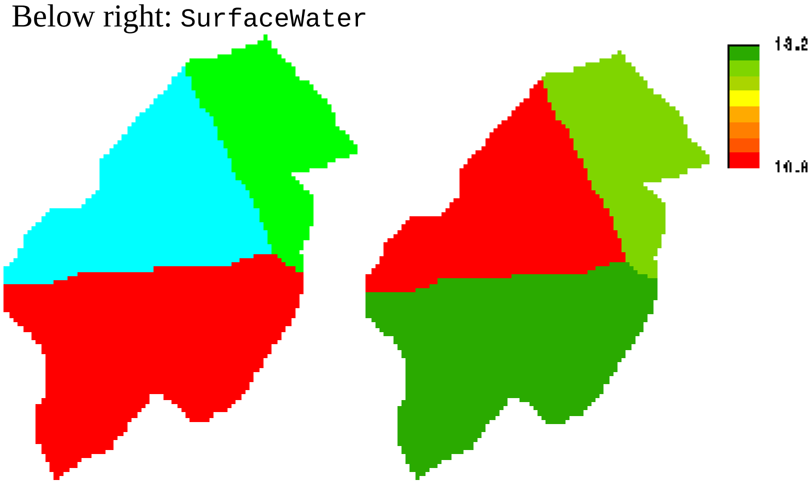
---

timeinputscalar

Timeinput functions are used to read temporal data. At each time step of the model an entry of a timeseries file or stack of maps is read. The example below reads a timeseries file at step 8 and assigns these values to a map based on `RainZones` map.

```
SurfaceWater = timeinputscalar(RainTimeSeries,RainZones);
```
contents of `RainTimeSeries` file:
```
.. first seven lines deleted
8 12.9 10.9 13.2
.. remainder of file deleted
```
Below left: `RainZones`

Below right: `SurfaceWater`



<<Return to article.

---

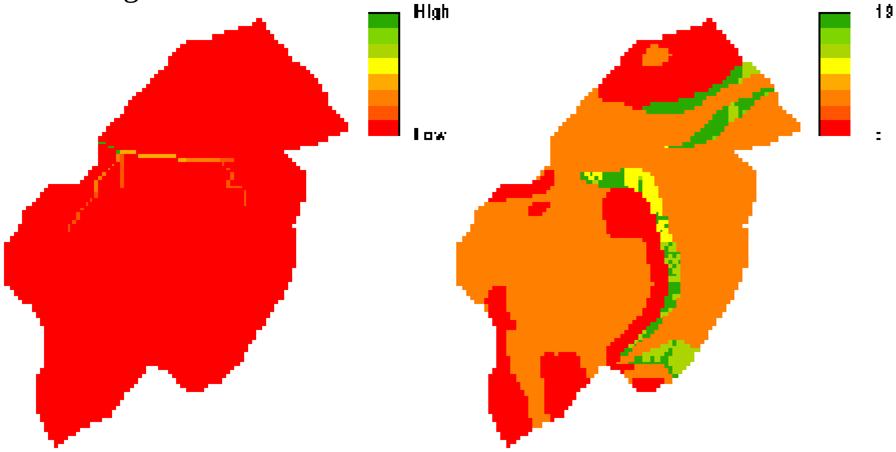
accuthresholdflux and accuthresholdstate

Accumulation functions are used to model the transport of material over a local drain direction network (`Ldd`). Different accumulation functions are available model transport under different conditions. Each accumulation operator can return two values: the flux map, which gives for each cell the amount of material that is transported downstream and the state map, which gives for each cell the amount that is stored in the cell. The accumulation function accuthreshold which is used in the example model transports material only when a certain threshold (in this case the infiltration capacity) is exceeded. The amount of water that actually infiltrates is given by the above mentioned state map (`Infiltration`) of the operator, the amount that runs of is given by the above mentioned flux map (`Runoff`).

```
Runoff, Infiltration =
accuthresholdflux,
accuthresholdstate(Ldd,SurfaceWater,InfiltrationCapacity);
```

Below left: `Runoff`

Below right: `Infiltration`



<<Return to article.

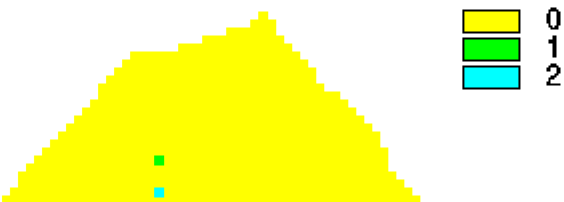
---

timeoutput

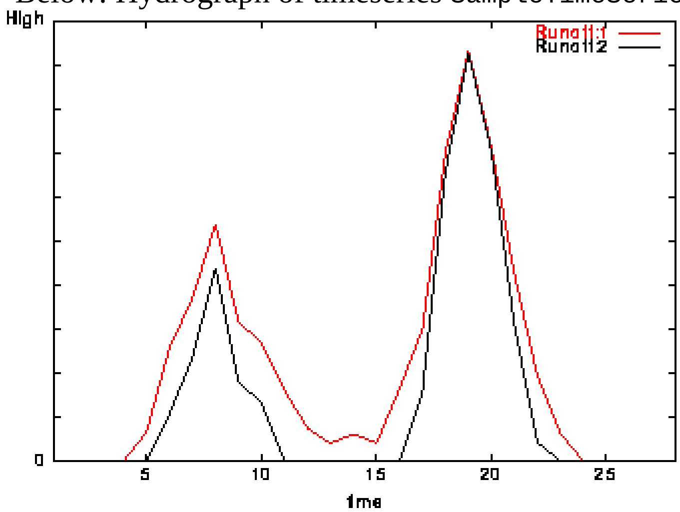
Timeoutput functions are used to sample the value of certain locations at each timestep. The values are stored in a timeseries file. In the example below two points are sampled, one point upstream of the sandy soil patch (red line) in the north and one point (black line) just downstream of the patch.

```
report SampleTimeSeries = timeoutput(SamplePlaces,Runoff);
```

Below: Top of map `SamplePlaces`



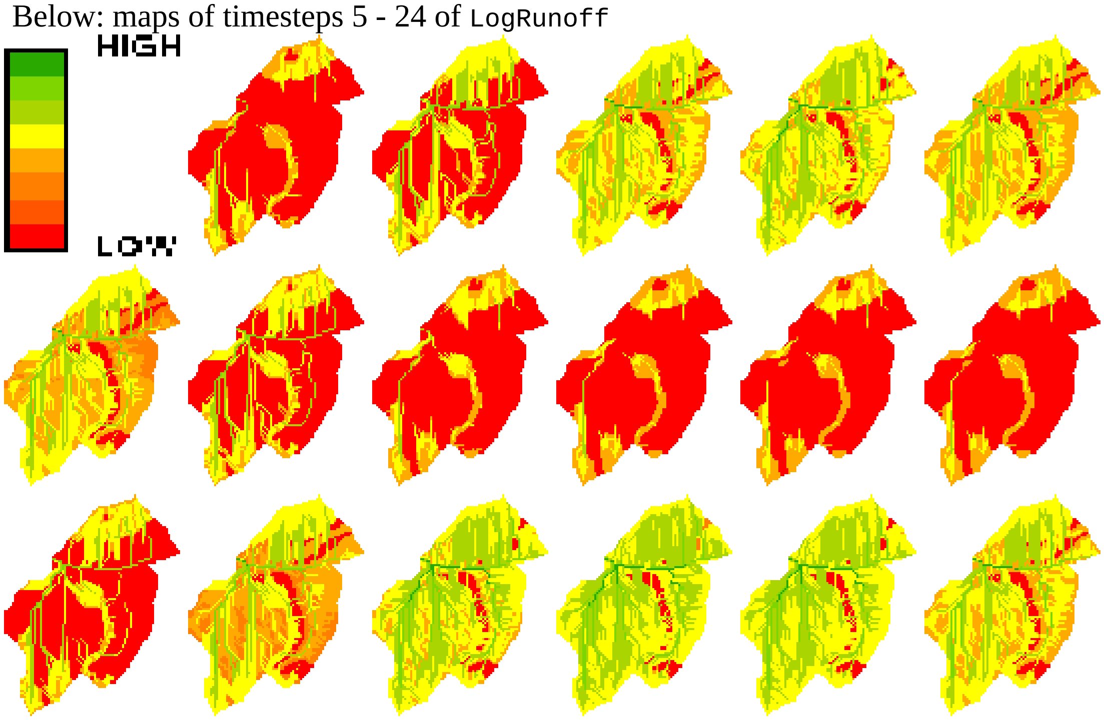
Below: Hydrograph of timeseries `SampleTimeSeries`

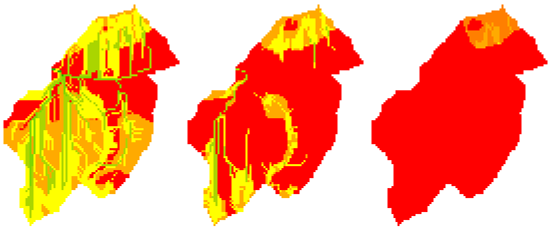


Spatial and temporal model results are stored in a stack of map layers. In the example below the logarithm of the runoff is written to a stack.

```
report LogRunoff = log10(Runoff+0.001);
```

Below: maps of timesteps 5 - 24 of `LogRunoff`





<<Return to article.

References

Aho, A.V., Sethi, R., Ullman, J.D. (1986) *Compilers: Principles, Techniques and Tools* Addison-Wesley Publishing Company.

Belmans, C., Wesseling, J.G. and Feddes, R.A. (1983) Simulation of the water balance of a cropped soil: SWATRE. *Journal of Hydrology* 63:271-286

Berry, J.K. (1987) Fundamental operations in computer-assisted map analysis. *International Journal of Geographic Information Systems* 2:119-136.

Burrough, P.A. (1996) Opportunities and limitations of GIS-based modeling of solute transport at the regional scale. *D.L. Corwin and K. Loague (eds) Applications of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone, Special SSSA Publication* in press.

Chow, V.T, Maidment, D.R. and Mays, L.W. (1988) *Applied Hydrology* McGraw-Hill.

De Roo, A.P.J., Jetten, J.G., Wesseling, C.G. and Ritsema, C.J. (1996) LISEM: A physically-based hydrological and soil erosion model incorporated in a GIS. *Applications of Geographic Information Systems in Hydrology and Water Resources Management HydroGIS 1996, IAHS Publication* in press.

Forrester, J.W. (1968) *Principles of systems. Text and workbook chapters 1 through 10.* Wright-Allen Press inc. Cambridge Massachusetts, USA.

Heuvelink, G.B.M. (1993) *Error propagation in quantitative spatial modelling applications in Geographical Information Systems* Doctor's dissertation, University of Utrecht, NGS 163.

Jenson, S.K. and Domingue, J.O. (1988) Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing* 11: 1593-1600.

Kemp, K.K. (1993) Managing spatial continuity for integrating environmental models with GIS. *Proceedings, Second International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modelling* Breckenridge, September 1993.

Kirkby, M.J. (1976) Hydrological Slope Models: The Influence of Climate. *E. Derbyshire Ed., Geomorphology and Climate. Wiley, London* pp: 247-267.

Leavesly, G.H., Restrepo, P.J., Stannard, L.G., Frankoski, L.A., and Santins, A.M. (1993) The Modular Modelling System MMS. *Proceedings, Second International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modelling* Breckenridge, September 1993.

Maidment, D.R. (1993) Environmental modeling with GIS. *Proceedings, Second International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modelling* Breckenridge, September 1993.

Moore, I.D., Turner, A.K., Wilson, J.P., Jenson, K. and Band, L.E. (1993) GIS and the land surface - subsurface process modelling. *Environmental modelling with GIS (ed. by M.F. Goodchild, B.O. Parks & L.T. Steyaert). Oxford University Press.* pp:196-230.

Pebesma, E.J. (1996a) *GSTAT, geostatistical modelling, prediction and simulation.* NGS 199.

Tomlin, C.D. (1990) *Geographic information systems and cartographic modelling.* N.J. USA: Prentice Hall.

Van Deursen, W.P.A and Heil, G.W (1993) Analysis of heathland dynamics using a spatial distributed GIS model. *Scripta Botanica* 21:17-28.

Van Deursen, W.P.A. and Kwadijk, J.C.J. (1993) RHINEFLOW: an integrated GIS water balance model for the river Rhine. *Applications of Geographic Information Systems in Hydrology and Water Resources Management, HydroGIS 1993 (ed. by K. Kovar & H.P. Nachtnebel), IAHS Publication No. 211* pp. 507-518.

Van Deursen, W.P.A. (1995) *Geographical Information Systems and Dynamic Models: development and application of a prototype spatial modelling language.* Doctor's dissertation, Utrecht University, NGS 190.

Wesseling, C.G. Heuvelink, G.B.M. (1993) *ADAM, an Error Propagation Tool for Geographical Information Systems* Cees G. Wesseling
Department of Physical Geography, Utrecht University
POBox 80.115
3508 TC Utrecht
The Netherlands
Telephone: (31)-(0)30-2532768
FAX: (31)-(0)30-2540604
E-mail C.Wesseling@frw.ruu.nl

Willem P.A. van Deursen
Resource Analysis
zuiderstraat 110
2611 SJ Delft

The Netherlands
Telephone: (31)-(0)15-2122622
FAX: (31)-(0)15-2124892
E-mail Willem.van.Deursen@resource.nl

Peter A. Burrough
Department of Physical Geography, Utrecht University
POBox 80.115
3508 TC Utrecht
The Netherlands
Telephone: (31)-(0)30-2532766
FAX: (31)-(0)30-2540604
E-mail P.Burrough@frw.ruu.nl

Robert G. Knox, Virginia Kalb, Elissa R. Levine, and Uday Bindingnavle

A Framework for Integrating Environmental Models to Simulate Forest Ecosystem Dynamics

ABSTRACT

FEDMOD (the Forest Ecosystem Dynamics Modeling Environment) is a modeling tool, developed using an object-oriented design, which allows multiple process models to be coupled in order to improve simulations of forest ecosystems. The system consists of a Graphical User Interface (GUI) which provides powerful, easy to use access to the configuration of various complex environmental models, and a simulation framework which orchestrates the simultaneous execution of and communication among the models. Using a generic query-response mechanism, parameter values from detailed models in one discipline are provided to drive models of other disciplines. In addition scaling parameters can be derived from detailed models which improve the accuracy or realism of models from other disciplines. Distinct reporting objects route information among models and process user requests for displays, logs, or summaries of simulated properties. An external clock object ensures synchronization among the models and reporting objects. Databases associated with each model instance save the state space at the end of each clock-step to provide a stable environment for queries. Prospective developments include adding more sophisticated management of spatial extent within the environment to allow model instances to have non-trivial forms of spatial overlap.

INTRODUCTION

A wide array of scientific and engineering disciplines develop environmental models. Although applied models may have narrow scopes and defined limited objectives, models developed to advance fundamental understanding may be used to explore a wider range of questions and to push the limits of present knowledge. For global change research, models are being asked to make predictions outside the range of present conditions. This precludes a strictly empirical approach, since only through a high level of realism in representation of critical processes is at all reasonable to expect valid extrapolations beyond available data (Hanninen 1995). Because ecosystems can be studied from many different perspectives, the best available representations of component processes come from an assortment of disciplines. For example, land surface models used in plant ecology or climate modeling employ highly simplified representations of soil hydrology, with one or two layers and aggregated parameters such as "rooting zone field capacity" or "profile moisture." When such a model, developed by one scientist or a small team, fails, it may be because of submodels that represent processes studied more rigorously by scientists from other disciplines (see

Bonan 1993, Bugmann and Martin 1995). Whether a more rigorous understanding would improve matters may remain an open question, because the primary model authors lack the disciplinary expertise and merging models from different disciplines would often produce excessively complex models (cf. Bugmann and Martin 1995).

For several years, we and our collaborators have been pursuing an alternative strategy for merging knowledge represented in different environmental models (Levine et al. 1993). Our aim has been to provide a flexible and reliable environment for simulation using multiple environmental models. We have developed an object-oriented framework called the Forest Ecosystem Dynamics Modeling environment (FEDMOD). Its primary purpose is to advance understanding by allowing simulation experiments, using communication among existing models that were developed by model authors from different science disciplines. Using this environment, a scientist can explore the effects of relying upon more rigorous representations of various ecosystem processes. They might then chose to maintain the present representations of non-focal processes--knowing that key features are consistent with a more specialized understanding, to adopt a more rigorous formulations, or to scale model coefficients to more closely approximate the behavior of the other model(s).

The FEDMOD environment encapsulates existing models to provide object- like behavior, rather than rewriting models in an object-oriented programming language. These models have a proven record of testing and use as stand-alone models of terrestrial ecosystems. As a design goal we strive to minimize changes to existing code, and we benchmark encapsulated models against the results of their stand-alone versions. In object oriented terminology, a particular encapsulated model then forms a "class," and a copy of a model running for a particular set of parameters is an "object" or "model instance."

A disadvantage of stand-alone models is the lack of synchronization on parallel runs. Without a centralized synchronization mechanism, it is impossible for models to dynamically share data. Thus, there is no simple and reliable way for scientists to create and study links among models. Also, types and formats for model outputs are pre-defined. Hence any change in the reporting of results implies a change in the source program of the model. During research use of these models there are frequent changes in the output required, and each change risks introducing errors to model algorithms. Also, when using a model from another discipline, a researcher has to carefully prepare model inputs ensuring that combinations of parameters are compatible and in the proper format. Even were parameters for a variety of locations available from a GIS, there are potential mistakes in formatting those data for use in each situation.

Hence, our goals were to create a reliable environment where these models could be run synchronously, where the researcher has the flexibility to select the output content and the type of reporting, and where less familiar models could be configured for a range of locations and run with some protection from errors in formatting parameters and checks for reasonableness on the values provided. We did not include interactive access to GIS data or real-time model linkages to a GIS. However, spatial data retrievals from a GIS have been used to populate simple databases of model parameters for common types of land units. These can be used interactively to configure model instances for situations of interest. This system addresses the range of issues that arise from the specialized, stand-alone nature of the original models. In the following, we describe the design philosophy, the operation of the simulation environment and the steps involved in adding an existing stand-alone model to the simulation environment.

DESIGN

After analyzing the requirements for such a system (see Levine et al. 1993), the design phase involved identifying the classes in the entire system and the types of relationships that exist between classes. In object-oriented software design (see Booch 1991), "classes" abstract the key features and required behavior of the things systems analysis identifies as making up the problem. Then, instances of classes, or "objects," contain data or represent states. Methods of a class (sometimes called member functions or valid operations on a class or object) define how other objects or people interact with objects of that class. All members of the class or of daughter classes derived from this class then support these same methods or operations. Thus for many circumstances, it isn't necessary to know exactly what class an object belongs to, since many objects share similar behavior. The relationship by which classes obtain some of their behavior from definitions higher up in the class hierarchy is called "inheritance."

The relations among some important classes are illustrated in Booch Notation (Booch 1991) by Figure 1. In FEDMOD, a class called "ecosystem process" abstracts the idea that the models we use represent the dynamics of a cluster of related features of a terrestrial ecosystem. Reporters service relationships between two models, or between a model and the user's visual system or an on-line file system. They define model outputs without the frequent changes to source code often characteristic of scientific programming. The also can define inputs to other models. The central design decision for this partitioning was to define all the flexible I/O as query operations among collaborating agents. By then associating a query or sequence of queries with an active agent, the reporter, any complexities of timing and message routing are hidden from client models that need the results of queries. The same query mechanism supports graphics routines that display how model states evolve. Once all encapsulated models are defined as ecosystem processes, operations such as queries can take the same form for any model. The rest of the software system was designed to provide a consistent environment to configure these agents and allow their behavior to evolve over time.

A second decision was to maintain the intuitive before-after ordering of time sequences and a metaphor between elapsed execution time and simulated time. This allowed us to be selective about recording sequences of state changes, but it also required the introduction of artificially discrete "events", during which time is synchronized and temporally consistent queries on arbitrary states are possible. For some existing models a fixed fundamental time step is consistent with their design. A clock tick comfortably maps on to a call or iteration of a loop. For other models, once encapsulated, they execute anonymously in between fixed time points when control is returned to the modeling environment.

Ecosystem process, Reporter, and Event are abstract classes that are only instantiated in their derived subclasses. (Thus, no instances or objects of class Ecosystem Process, Reporter, or Event ever exist--just objects of classes that inherit from these classes.) The three classes follow directly from the decisions just described. Several other key classes were needed to manage them in a particular distributed computing system, and to support the metaphor of a physical place or landscape in which these agents interact.

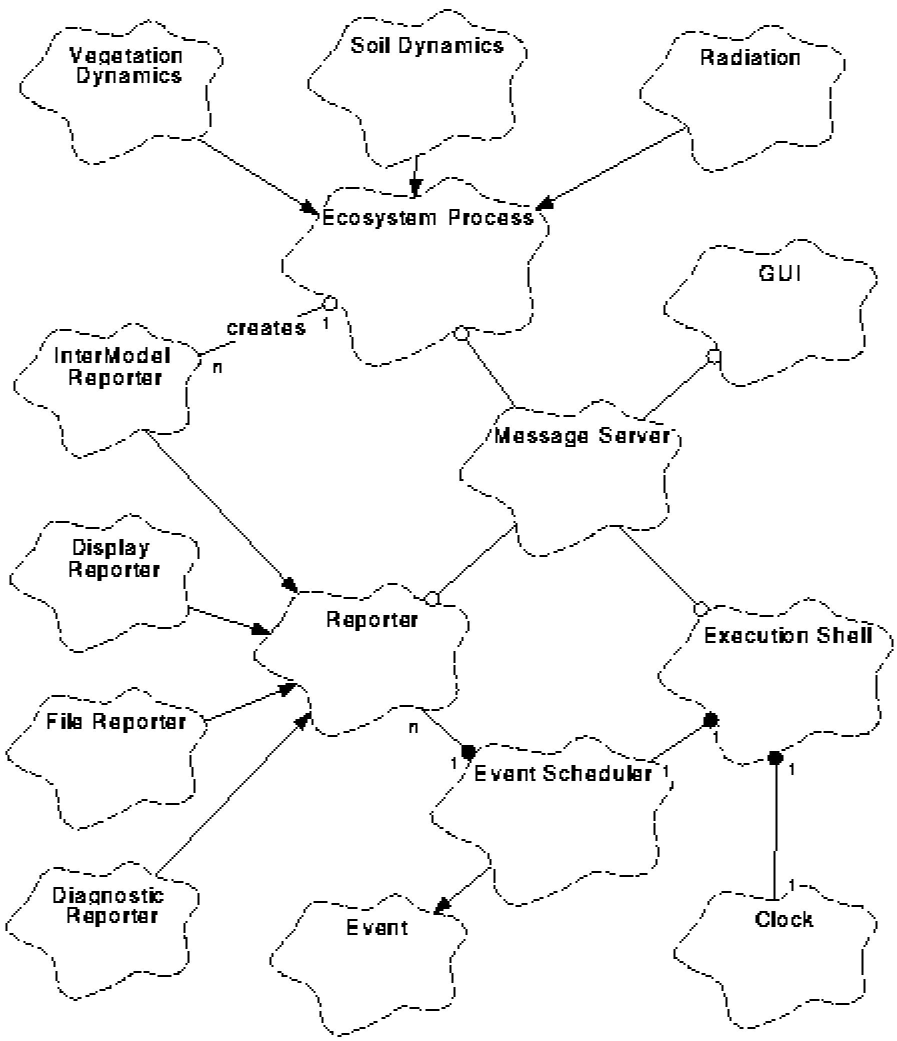


Figure 1. Class Diagram in Booch Notation. Arrows show class inheritance, pointing to the parent class of a class that inherits from another class. Lines with filled circles show which classes contain instances of other classes. Lines with open circles show classes that use other classes to implement their operations or methods.

Abstract Classes

The Ecosystem Process class captures the common characteristics shared by every ecosystem model in the FEDMOD environment. Some of these include the existence of a special initialization routine or routines, a model step value, which is the length of a distinct simulation time-step for the model, and a way of accessing the model's attribute values. Attributes are the model state variables, input data, and simulation results that the encapsulated model makes available for query. Attributes are defined in consultation with scientists intimately familiar with the model. They include the contents of variables useful to diagnose problems or understand model behavior, as well as variables normally output by the stand-alone model. When the environment is able to create a place in the computer's file system for a model to run, and confirms that the program to be executed is a valid encapsulated model and that the initialization procedure was successful, then the instance of an ecosystem process is said to be alive and ready-to-run. The common behaviors for each encapsulated simulation model are represented as an inheritance from the Ecosystem Process abstract class. Thus an inherited shared definition forms the basis for how simulation models appear in the environment.

Each model is different, in that configuration requirements, required parameters, and simulated attributes are all different. The common behavior and interface definition provided by the abstract class allow these differences to be expressed in consistent ways.

The Reporter class maintains the common characteristics that exist in each type of reporter. Reporters are distinct agents that are responsible for querying the simulation models at a specified reporting frequency and presenting the output in the manner requested. The manner of presentation is a characteristic of the type of reporter. To date, we have identified four types of reporters: Display reporters that provide a continuous graphical output in the specified graph form such as Stripcharts, Scattergrams, or Connected-line graphs; File Reporters that are responsible for dumping the results of the queries into a specified output file; Diagnostic Reporters that query model attributes that might not otherwise be queryable; and Inter-Model Reporters that provide a means of accessing the attribute values of a running model for use in another running model. The common characteristics for each reporter type include the "name" of the attribute that needs to be queried, the Model "instance" to which the attribute belongs, the "steps" which indicates the frequency of reporting, and the "query type" which specifies that either Point Query, Averaged Query, Summed by steps, or Cumulative Summing needs to be performed. An important application of reporters is to show the interactions among model attributes during simulations, using one of the types of Display Reporters.

The Event class provides a generic set of methods used to schedule discrete events. This class maintains the current simulation time value and keeps the system informed when the time value changes (increments, resets, etc.). It provides a "Time" attribute, which reports the current time value, for use by the events that are scheduled by the derived classes of this abstract class. It also contains an internal flag called "Time_Changed", which is set after the scheduling of events has been completed for the current time step. Finally, there are two attributes called "Step" and "StepLen", which save the step-value and step-length parts of the simulation time. The actual simulation time in seconds can be obtained by multiplying the "Step" value by its units ("StepLen"). Splitting simulated time into two attributes helps avoid numerical overflows likely if century long simulations were measured solely in integer seconds. This class also contains access functions to access and update the attributes.

Classes To Manage Collaborative Computation

The Event Scheduler class schedules reporters based on the clock step. There is a single instance of this class, the Event Scheduler object, for each ecosystem simulation that is active. As reporter objects are created, they get registered by the Event Scheduler object, which later activates the reporters that need to run at any particular clock step. The Event Scheduler object is driven by the clock object. At each transition of the clock step, the Event Scheduler object checks to see if any reporters need to become active at that step. The Event Scheduler class contains "Add_Reporter" and "Del_Reporter" methods, which respectively register and remove reporters from the Event Scheduler object, a "Reporters2Activate" method that generates a list of reporters to run at a given clock step, and an "Activate_Reporters" function that activates the reporters from the list created by the Reporters2Activate method. A singly linked linear list is used to store the list of reports.

The Clock class maintains the overall simulation time information. When a clock object is created, it negotiates a meaningful clock step based on all the model and reporter interval steps. To avoid arithmetic overflow, the simulation time is represented by a value part and a step-length part. The units for step-length are seconds. The clock object is responsible for broadcasting the new time after each clock increment, to each model that is being simulated. There is a provision to introduce a user-specified delay between each clock increment. This delay can be specified through the user interface, and is useful when the simulation speed needs to be slowed down in order to study the results that are continuously displayed by the display reporters.

A system clock and facilities for scheduling events are a normal part of a high-level simulation language. We have re-created them for use with existing simulation models, written in procedural languages.

The Message Server class provides a means of communication between model processes and reporters. There are two levels of communication:

1. the communication between the GUI (clients) and the simulation environment (server), which is implemented using remote procedure calls (RPC) (UNIX 1990a) and,
2. the communication between the various processes within the simulation environment (including the encapsulated models, reporter objects, clock and the event scheduler object), which is implemented using standard Unix System V message queues (UNIX 1990b). By providing this as a distinct utility class, we can more easily move to new communication protocols for ways of further distributing the computation across processors or workstations.

The Execution Shell class creates model processes, starts up a clock instance and an Event Scheduler instance, registers reporter objects to the Event Scheduler object, communicates with the GUI, and ensures that all the models are synchronized with respect to the clock object. It represents the overall simulation arena or framework.

USING THE SIMULATION ENVIRONMENT

The RPC Server

An instance of the RPC Server should always be running in the background on any compute server (remote host machine) supporting FEMOD. This server contains the messaging functions that enable communication between a GUI and the part of simulation environment where the models are run. In addition to a master server that routes requests from remote machines, each scientist using the FEDMOD environment starts up a server process using their own account. This user process then handles authentication of remote requests and owns all of the simulation processes and files created while running that person's simulations. By distributing the environment in this way, the graphical interface (GUI) can be run locally while the simulations are run remotely on the fastest machine available.

User Interface

The Graphical User Interface (GUI) is a separate set of programs from the actual models and the classes that manage collaborative computation. This permits the compute-intensive models to be resident on a powerful compute server, and permits the GUI to provide quick response time by running as a local process. The GUI was developed with TAE+ (Szczur and Sheppard 1993), a Motif Interface builder created under contract to NASA. At the top level, the GUI provides the overall management of the modeling session. The definition of overall parameters (e.g., simulation length, spatial extent), choice of specific models for simulation, and run-time control are handled through a menu bar that is visible throughout the simulation session.

Each model has its unique requirements for configuration. Hence, the selection of a model type triggers a model-level panel hierarchy which has been customized for each type of model. Although the parameters are specific to one type of model, much of the code that defines the user interface for each model is shared. Thus, models encapsulated in FEDMOD have a common appearance and it is relatively easy provide interfaces for new models.

Specifying Simulation Models

In the main panel, one can select the "Models" item from the menu bar. This results in a pull-down menu with "Framework", each of the encapsulated models available, and "Configured Models" as the menu items. Initially, the only accessible choice is Framework, since the global modeling parameters must be specified first. After successfully completing the top-level specification panel, the model-specific menu items become accessible. A particular model can now be selected. Although each model must be selected and configured separately, the researcher is free to select multiple models for a single simulation session, including multiple instances of a particular model. The Configured Models menu selection shows any models have been selected, and their configuration status (e.g., ready-to-run, or awaiting user configuration). Since configuration can start from either default parameters or a previously saved configuration, it is easy to set a series of model instances with similar parameters.

Configuring Selected Models

Each type of encapsulated model has a distinct sub-interface consisting of a model-specific main panel and panels for configuring instances of that model and any associated reporters. When a model type is selected, the appropriate sub-interface becomes active. Configuration panels are specialized to provide access to parameters that define how model instances differ when run for different sites or under alternative scenarios. Figure 2 illustrates an initialization panel for the Hybrid v3.4 model (see Friend et al. 1993; Friend et al. in press).

Values entered with the GUI are checked against a series of rules or constraints to insure that they represent part of a valid set of run parameters for this type of model. Rather than being coded into the GUI, these constraint sets are tables that researchers can edit to define more or less restrictive conditions for simulations with a particular model. By selecting among different constraint sets, researchers using the environment could have relatively restrictive, safe choices for routine use or for using unfamiliar models, or permissive environments for testing the limits of familiar models. Where constraints are used define how the GUI looks, for example upper and lower limits for a slider widget, the constraints are supplied each time a new model is selected. Thus, the sub-interface directly reflects how permissive the current rules are. Violations of rules are reported with alert panels that pop up when a violation is detected, but the researcher may chose to ignore the rule by over-riding the warning. Similarly, key-in fields are available to enter values outside the end points of a slider widget or to easily enter specific numerical values that may be awkward to select with a graphical widget.

The present version of FEDMOD has simple constraints on single parameters and pairs of parameters implemented in a generic way. More complex constraints are either handled by custom code for a particular model, or by grouping the parameters in external files that are selectable as parameter sets but not specified individually with the GUI. An important change to improve robustness would be to further consolidate definition of the parameter lists for each model. Multiple copies of these lists embedded within the rule-checking classes and the classes that format parameter files to be read by models are a potential source of maintenance problems. They also increase the effort required to add new models by reducing the fraction of code that can be reused.

After the configuration is defined, selecting "Add Model" from a pull-down menu under the left-most "File" menu bar item generates a request to the Execution Shell to create and initialize an instance of the model using the specified parameters. Success of the Add_Model operation means that the model instance exists and is ready-to-run. A partial configuration will also persist, so the full configuration needn't be completed before starting work setting up other model instances.

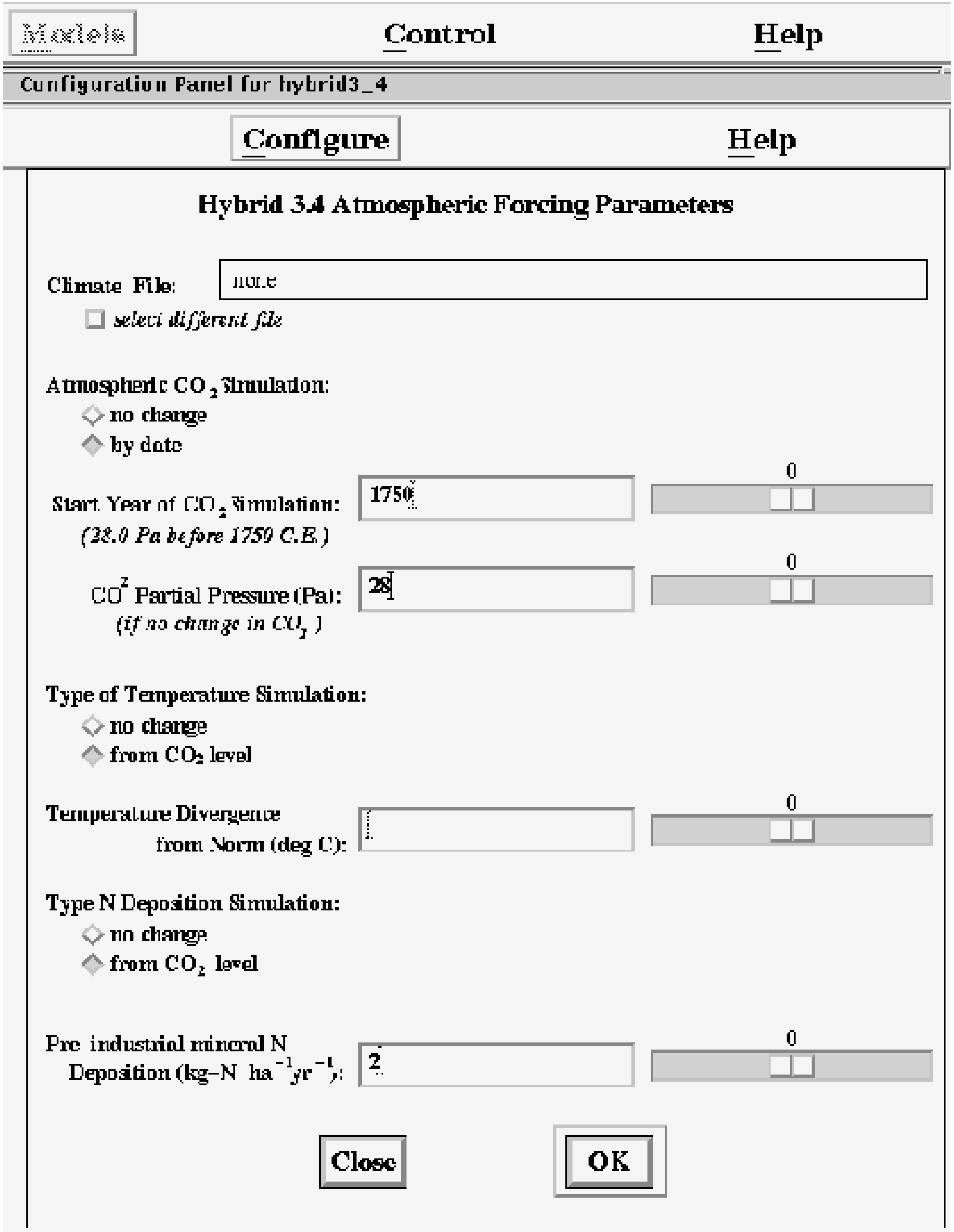


Figure 2. Sub-Interface for Hybrid v3.4 with a Configuration Panel.

Specifying Reporters

Under the model-specific panel or sub-interface, reporters are added by selecting the VIEW menu bar item and then the type of reporter in the resulting pull down menu. Based on the type of reporter selected, the GUI requests pertinent information from the user and sends the information to the simulation environment (Execution Shell) through the message server. The

Execution Shell then creates the Reporter instances and registers them to the Event Scheduler object (using the Add_Reporter method). Reporters that combine information from different model instances, including different types of models, may be created using the "View" item on the main menu bar.

States of the Simulation

The simulation has five distinct control commands that initiate changes in the overall state of the simulation environment. The commands are: Run, Halt, Suspend, Resume and Reset, and the four states are: running, halted, suspended and reset. The running state means that simulation computations are underway. The halted state indicates that processing has stopped and cannot readily be restarted from that point. The suspended state represents a paused simulation that preserves a snapshot of the simulation prior to the Suspend command. Normally this is implemented by stopping the clock object and leaving model processes and reporters waiting for the next clock tick (see below). Reporters can be added to a suspended modeling environment, but changes to configured models that might require reinitialization are prohibited. This permits a researcher to add a graphical display or log another state variable to learn more about a phenomenon that has just appeared in the simulation. The Resume command follows a Suspend, and implies that models continue from the snapshot prior to the Suspend. After a Resume, the simulation returns to a running state. The Reset command must act on a suspended simulation and differs from a Resume in that, it disregards the snapshot of the simulation prior to the Suspend and reinitializes the simulation by resetting the clock to the first time step and reinitializing the active models. It then expects a Run or Halt command to advance to the next state.

Starting the Simulation

Once the researcher has specified the model(s) to be simulated and any reporters for the model(s), she uses the Control menu bar item from the main panel to issue the Run command, which starts the simulation. (See Figure 3.) When the GUI routes the Run command to the Execution Shell, the execution shell creates a clock instance, with the clock step being a value negotiated between the step values of all the models and reporters specified. (The negotiated value is the largest integer step length that will insure that all the objects in the environment can be run as often as requested, so sets of step lengths with small greatest common divisors, e.g., 1, should be avoided.)

For each clock increment, the clock object broadcasts the time value to the Event Scheduler object and each of the ecosystem processes. For each clock increment, the Event Scheduler object activates reporters scheduled to run at that time step. The reporters query a states database (see below) for one or more model instances and report back the result in a way characteristic of the type of reporter. After the reporters have been activated, the models evolve to their next states and the states database is updated. Hence at any time step, the reporters query the states database that reflects the stable state for the previous state evolved. In case of display reporters, the results of the reporter queries are continuously plotted on a graph. File reporters log query results to a disk file, and so on. Note that for efficiency reasons, reporters were grouped under a centralized scheduler, whereas model instances are more autonomous--receiving the broadcast time value directly.

Figure 3. Main GUI panel with the Run command selected under Control on the menu bar.

Suspending a Simulation

Once the simulation is running, it may be paused at the end of any time step by issuing the Suspend control command. When this command is sent to the simulation environment, the clock object, which constantly polls for the existence of a Suspend control command prior to each clock increment, goes into a wait state. All the processes are thus put to sleep once they complete any on-going calculations. The clock is triggered out of the wait state by the arrival of a Resume control command from the GUI. If a Halt control command is issued instead, then the clock object considers this to be the end of the simulation and broadcasts a special code to all the model processes. The researcher can add or modify reporters while the simulation is in a suspended state, but not model instances.

ENCAPSULATING A STAND-ALONE MODEL

To encapsulate an existing stand-alone ecosystem model, apart from providing a model-specific GUI and the other services described above, a series of changes are made to the model code itself and to specialize the ecosystem process class to support this type of model. We designed these procedures to be relatively simple and easy to repeat when new model versions are provided by the models' primary authors. They include making some limited changes to the existing model code and providing new code to supply responses to queries and other required behavior.

Modify a Stand-Alone Model

The stand-alone model is revised so that it can be invoked (started up) by a new external procedure. Thus any main module, which in several languages is the starting point of execution in a program, must be converted to a callable unit, such as a subroutine. The second important change is to prepare the stand-alone model to accept an external time token, which would be used to evolve the model to its next state of simulation. In stand-alone programs, time is often simulated with some kind of an iterative loop, with the maximum clock time signifying the last iteration of main loop. Loops that serve to step the model states through time must be converted to conditional blocks of code that are executed whenever the time token meets certain conditions.

The steps involved in modularizing the stand-alone model are as follows:

1. Replace the main module by a procedure with two arguments. The first argument is the global clock value (passed by the simulation environment) and the second argument is

the ModelStep value which provides the required frequency of invocation for model code. This value is set by the encapsulated model during an initialization phase.

2. Since the original main program is now a procedure, in some languages all the variables declared within this procedure are now considered local. So, when the procedure exits and returns to the calling program (the simulation environment), the contents of these variables are lost. To prevent this from happening, variables previously declared in the main module must be converted to either global or static variables. In Fortran, for example, variables require the SAVE attribute.

The process of preparing the stand-alone model to evolve its states in response to an external clock involves identifying all the clock emulating loops in the stand-alone program, and replacing them by decision structure (IF--END IF) that ensure the correct frequency and timing. For example:

```
FOR Hour = 1 TO 24 DO
        :
        :
END DO
```

Figure 4: Pseudo-code representing hourly calculations with a loop.

```
/* Hourly interval in seconds */
IF (Modulus (Clock, 3600) = 0) THEN
        :
        :
END IF
```

Figure 5: A decision structure representing hourly calculations, for a fixed clock step length of one second.

As implemented, the timing code is slightly more complicated than this, since clock ticks will represent a negotiated step length rather than integer seconds. The divisor in the modulus operation will be calculated from the ratio of the model's intrinsic frequency and the external step length. For example, if the global step length is 3600 seconds (one hour) the clock will be in units of one hour, and a model running at an hourly step will take the modulus of Clock and 24 for daily operations. For code blocks executed once for each call at the requested frequency, the conditional statements can be omitted, assuming the calling frequency is correct. Some care may be required to control execution of sections specifically assigned to the beginning or end of a longer time interval, say the first or last hour of the day.

Creating A Model Class

A Class representing the new model inherits from the abstract class: Ecosystem Process. This new class must encapsulate any specific characteristics of the new ecosystem model. The model interfaces (methods/valid operations) are designed so they are consistent regardless of type of ecosystem model. An important departure from the usual examples of object-oriented software design is that the names of model attributes or state variables are treated as qualifiers to general query methods, rather than supporting a distinct access method for each attribute.

Experience has shown that the list of supported attributes is one of the more frequent things to change during model revisions. Authors of ecosystem models often change the formulation of minor processes, altering the way states are partitioned. By treating the attribute name as a modifier to a generic query, we can avoid changes to the model's formal interface, while tracking substantive changes in the underlying model.

Inter-Language Calls

The stand-alone ecosystem model could be implemented using any major computer programming language. In order to communicate with its stand-alone code, we needed interlanguage function calls. (The simulation environment is implemented using C++). The bare minimal set of interlanguage functions includes: **CInitialize, CAttrCount, CAttrNames and CGetAttr**. The **CInitialize** function invokes the initialization part of the ecosystem code. The **CAttrCount** function obtains the number of queryable attributes from the ecosystem model, while the **CAttrNames** function obtains the list of supported attribute names. The **CGetAttr** function gets the current value of a specified attribute from the states database. All these interlanguage functions are grouped into one location (a file) in order to:

- Avoid a tight-coupling between the model interface (the C++ model class) and the encapsulated model code, so that the Model Interface remains unaffected by changes in the encapsulated model, and
- Group the interlanguage functions in one location so as to make them easier to maintain. The interlanguage features are compiler-dependent and hence may need changes when new versions of compilers are used or when porting to a different operating system.

Ideally, for any particular source language for ecosystem models and set of compilers, there should be one definition of the interlanguage calling functions. For historical reasons, FEDMOD has versions associated with each of the main types of model encapsulated to date. We plan to consolidate these when developing a version of the environment to distribute to other researchers.

Model-Specific States Database

A model-specific states database is implemented to store the stable state values of the model attributes at the end of each complete time step. The only requirement for the states database is to store the model attribute values in a queryable form. This could be implemented using various data structures. While selecting the data structure, it is important to consider the size of the attribute set, the requirement that it be easily accessible (to read and write) with low execution overhead, and that it needs to be hidden from other parts of the environment to avoid erroneous or accidental changes. If Fortran 90 is being used, we recommend the use of **modules** (Metcalf 1992) to represent the states database. Along with a states database, we implement access functions that get and set attribute values.

Although it adds some execution overhead, we use string matching on attribute names to resolve accesses to particular attributes. This permits use of generic access methods (one for each of the intrinsic types in use) that are only resolved to specific attributes during execution, even when the language of the original model does not support run-time binding. At present, although access methods are fully generic C++ functions, some dependence on the list of attributes extends through the C++ and interlanguage layers. Restricting dependence on the actual list of attributes to a single module, that would also supply the master list to other procedures, would make this part of the system easier to update and maintain.

CONCLUSION

The FEDMOD simulation environment enables us to perform simulations of several instances of different models in parallel, with the flexibility to dynamically select which models to run and which attributes to report and display. Models for forest tree population dynamics, soil physics, snow physics, terrestrial biogeochemistry, and physics of canopy thermal regimes have been encapsulated for use in the environment. We have several series of simulation experiments planned or underway, which use intermodel reporters to share information among models from different disciplines. As expected, new models have become progressively easier to encapsulate and include in FEDMOD. Also, we find areas for improvement and as other model authors release new versions, we are able to incorporate them readily--with substantially less effort than the initial model encapsulation.

Experience encapsulating these models has identified parts of the initial design where an even greater emphasis on code reuse and strict object-oriented design would improve the robustness of the FEMOD system. We are also exploring adding routines to support client models using results of queries to scale calculations from internal functions rather than directly using a queried value from the previous time step. Another aspect where work is in progress is in extending the design of the model specification to include spatial extent. With this feature, we would be able to differentiate instances of a model based on their extents and take into account non-trivial spatial overlaps among model instances. Intermodel reporters would then return an appropriately weighted average from all the overlapping instances.

ACKNOWLEDGEMENTS

We thank K. Hammond for help in the "Encapsulating a Stand-Alone Model" section of this paper, and D. Kendig for work encapsulating additional models and streamlining the modeling environment. Funding for this work was provided by the Office of Mission to Planet Earth, U.S. National Aeronautics and Space Administration.

REFERENCES

Bonan, G.B. (1993) Do biophysics and physiology matter in ecosystem models? *Climatic Change* 24: 281-285.

Booch, G. (1991) *Object Oriented Design with Applications*. Redwood City, CA: Benjamin/Cummings.

Bugmann, H. and Martin, P. (1995) How physics and biology matter in forest gap models. *Climatic Change* 29: 251-257.

Friend, A.D., Shugart, H.H., and Running, S.W. (1993) A physiology-based gap model of forest dynamics. *Ecology* 74: 792-797.

Friend, A.D., Stevens, A.K., Knox, R.G., and Cannell, M.G.R. A process-based biogeochemical, terrestrial biosphere model of ecosystem dynamics. *Ecological Modelling* (in press).

Hanninen, H. (1995) Assessing ecological implications of climatic change: Can we rely on our simulation models? *Climatic Change* 31: 1-4.

Levine, E., Ranson, K.J., Smith, J.A., Williams, D.L., Knox, R.G., Shugart, H.H., Urban, D.L., and Lawrence, W.T. (1993) Forest ecosystem dynamics: linking forest succession, soil process and radiation models. *Ecological Modelling* 65: 199-219.

Metcalf, M. and Reid, J. (1992) *Fortran 90 Explained.* New York: Oxford University Press.

Szczur M.R. and Sheppard, S.B. (1993) TAE Plus: Transportable Applications Environment Plus: A User Interface Development Environment. *ACM Transactions on Information Systems* 11(1): 76-101.

UNIX System Laboratories, Inc. (1990) *UNIX SYSTEM V RELEASE 4. Programmers' Guide: Networking Interfaces*. Englewood Cliffs, NJ: Prentice Hall.

UNIX System Laboratories, Inc. (1990) *UNIX SYSTEM V RELEASE 4. Programmers' Reference Manual*. Englewood Cliffs, NJ: Prentice Hall

---

Robert G. Knox(a), Virginia Kalb(b), Elissa R. Levine(a), and Uday Bindingnavle(c)

(a) NASA/Goddard Space Flight Center, Biospheric Sciences Branch, Code 923, Greenbelt, MD 20771

(b) NASA/Goddard Space Flight Center, Laboratory for Terrestrial Physics, Code 920.2, Greenbelt, MD 20771

(c) Hughes STX Corporation, 7701 Greenbelt Rd, Ste.397, Greenbelt, MD 20770

James D. Westervelt
Lewis D. Hopkins

Facilitating Mobile Objects within the Context of Simulated Landscape Processes

This effort develops and demonstrates fundamental approaches for the simulation of mobile entities within the context of dynamic landscapes. As computing power continues to become cheaper and faster, Geographical Modeling Systems (GMS) will become increasingly important for the intelligent management of landscapes. However, a number of technical challenges must be met before GMS capabilities are widely accepted, two of which are addressed by this research. First, the ability to simulate the behavior of individuals is currently lacking in landscape decision support systems. Population modeling is unsuitable for very low population densities where the locations of individuals on a large, diverse, and fragmented terrain become important. Second, because landscape processes occur and are modeled at a variety of spatial and temporal scales, it is necessary to allow simultaneous simulation of disparate scales. In a management setting , techniques for linking distinct landscape models to run simultaneously are necessary. All of the capabilities should run on a common platform with a consistent user interface. More importantly, each submodel can relax the typical requirement that the landscape state be constant and instead access dynamically varying system state information from concurrently running models.

Background

The United States Army is responsible for the management of military training areas across the United States and the world. To help manage this land the Army Corps of Engineers' Construction Engineering Research Laboratories (CERL) have conducted research, design, and development of computer based modeling and simulation environments. Examples include GRASS (Westervelt, et. al. 1992)), surface interpolation algorithms and techniques (Mitasova, et. al, 1994), and overland and river flow simulation (Saghafian, 1993; Gaur & Vieux, 1992; Vieux and Westervelt, 1992, and Rewerts & Engel, 1991). The Geographic Modeling Systems Lab (GMSlab), located at the University of Illinois at Urbana-Champaign, in conjunction with CERL, is currently pursuing a program to extend the simulation capabilities and to link together multiple landscape simulation software packages into an Integrated Spatio-Temporal Ecological Modeling System (I-STEMS). The research reported here extends the general purpose simulation tools and links two different simulation capabilities together. A discrete entity based simulation environment was designed and developed to extend the available landscape simulation tools. That environment was then linked to GRASS, which provided a static landscape for the entities, and Maxwell's (1995) Spatial Modeling Environment (SME), which provided a dynamic landscape option.

Entity Design

A number of requirements were established for the design of an entity based simulation environment. The design focuses on capturing the behavior of individual animal objects, which update their internal state using time steps in the range of a day to a month. These animals must be able to interact with landscape information as provided in raster format from a GIS or from a dynamic landscape simulation running concurrently with the animal simulations. Animals must be able to locate and evaluate other mobile animal entities in a dynamically changing home range (established for prototype purposes as circular). Once evaluated, the animal objects must be able to interact with one another both synchronously and asynchronously. Figure 1 shows how the functionality of the animal objects was divided into a number of distinct object classes which were then programmed using the IMPORT/DOME object-oriented simulation language (Morrison, 1995). Note first that the development adhered to a software development approach that separates the actual software model from the software that controls and visualizes the model. Within the model classes are classes that provide interactions between animals, between animals and the landscape, individuals and populations, and between classes that define an animals internal interactions. Figure 2 shows how animal classes (bottom of the figure) can use any one of a number of prototypical animal classes as a superclass, modified to meet the specifications of each animal. In turn, these superclasses utilize a variety of different animal internal classes, which provide the primary instructions for updating the animal's state and location. All animal superclasses themselves utilize an AnimalInfo superclass which provides a means for sharing information between the various classes.

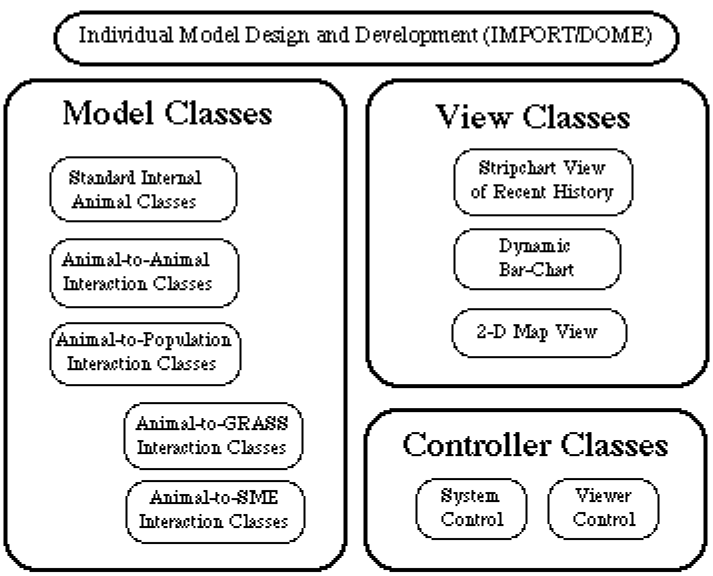


Figure 1: Entity Design Approach

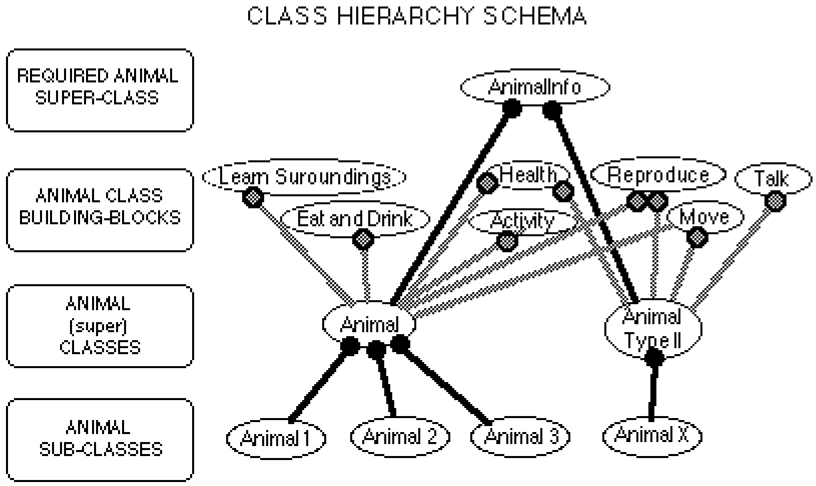


Figure 2: Class Hierarchy

Integration

Animals interact with landscapes as well. To provide landscapes a pair of IMPORT/DOME MapInfo classes were developed. One encapsulated standard GRASS software libraries, which allowed direct access to landscape information in its native GRASS format during a simulation run. The other encapsulated an SME generated landscape simulation model based on a unit-model designed and developed using the Stella simulation environment. Figure 3 depicts a portion of a Stella model that was used for the demonstrations. This landscape model is part of a Desert Tortoise habitat suitability model developed by the author and others (Westervelt, et. al, in press).

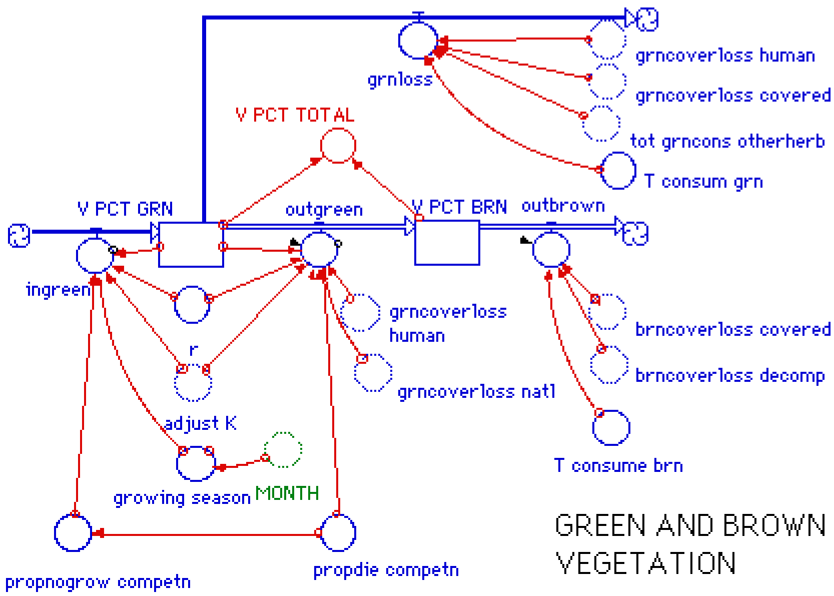


Figure 3: Landscape Model (portion) in Stella

The dynamic and static classes were completely interchangeable allowing a rapid transition from a static to a dynamic landscape. The static landscape is preferred for the development in animal classes because 1) simulations run much faster and 2) the number of changing variables is reduced. The integration of the components leading to dynamic and static landscape simulations is depicted in Figure 4.

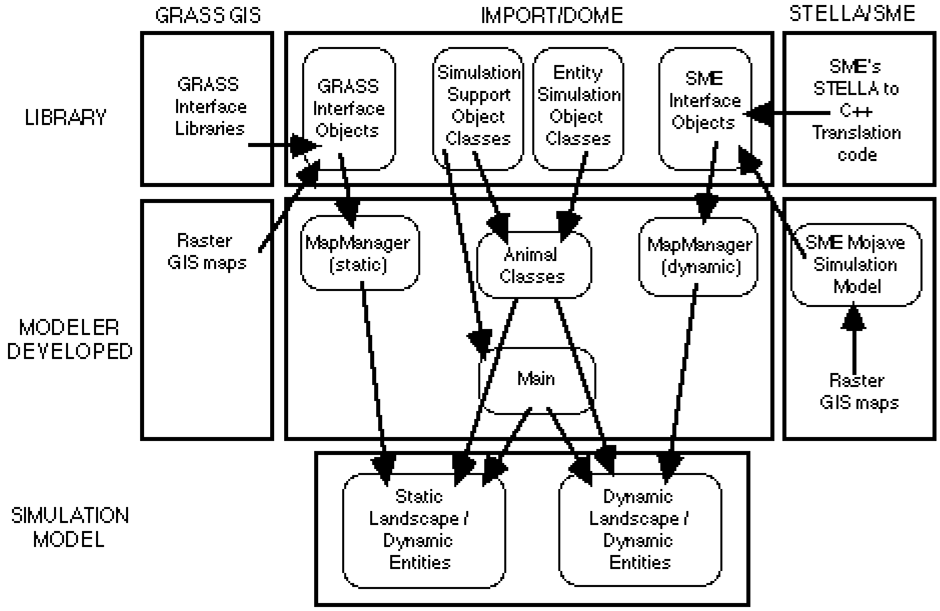


Figure 4: Static and Dynamic Landscape Integration

View Ports

A number of viewports were developed to allow the modeler to view current and historical information. First, a map viewport will be seen in the demonstration graphics. This provides a running picture of the current location of the entities being simulated. As a backdrop an appropriate static picture is provided; for the demonstrations this is a shade-relief map image of the test area. Second, to provide current simulation views of the state of any individual entity and it s surroundings, a barchart viewer was developed (Figure 5). Finally, a generic StripChart class was created to allow for the plotting of historical information about some aspect of the simulation. Figure 6 shows a sample trace of temperature and rainfall.

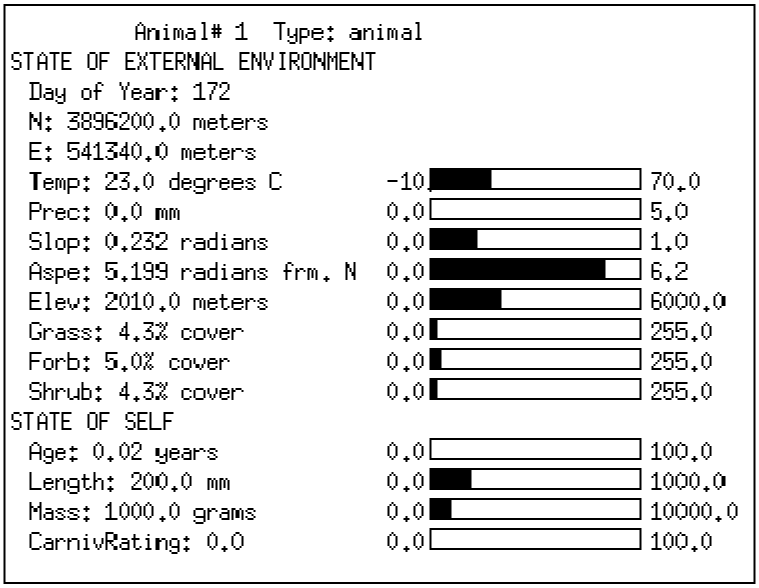


Figure 5: Bar Chart Viewport

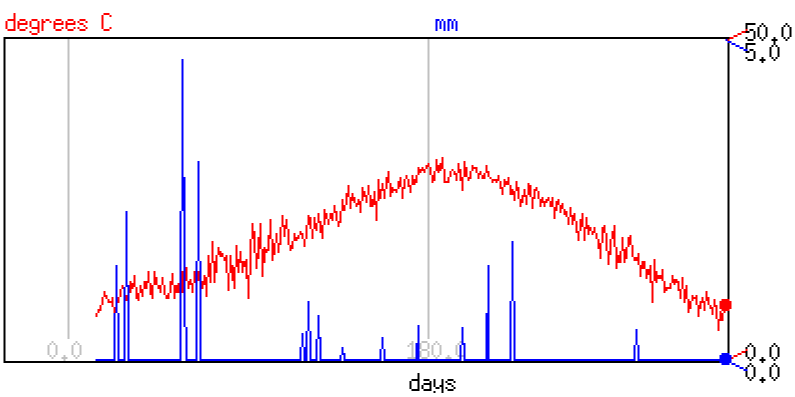


Figure 6: Strip Chart Viewport

The overall structure of the object-oriented classes developed and linked together is depicted in Figure 7. At the base of the figure is the "main" class which , at run-time, instantiates (brings into existence) objects based on the classes immediately above it, which in turn instantitiate objects based on the classes above.

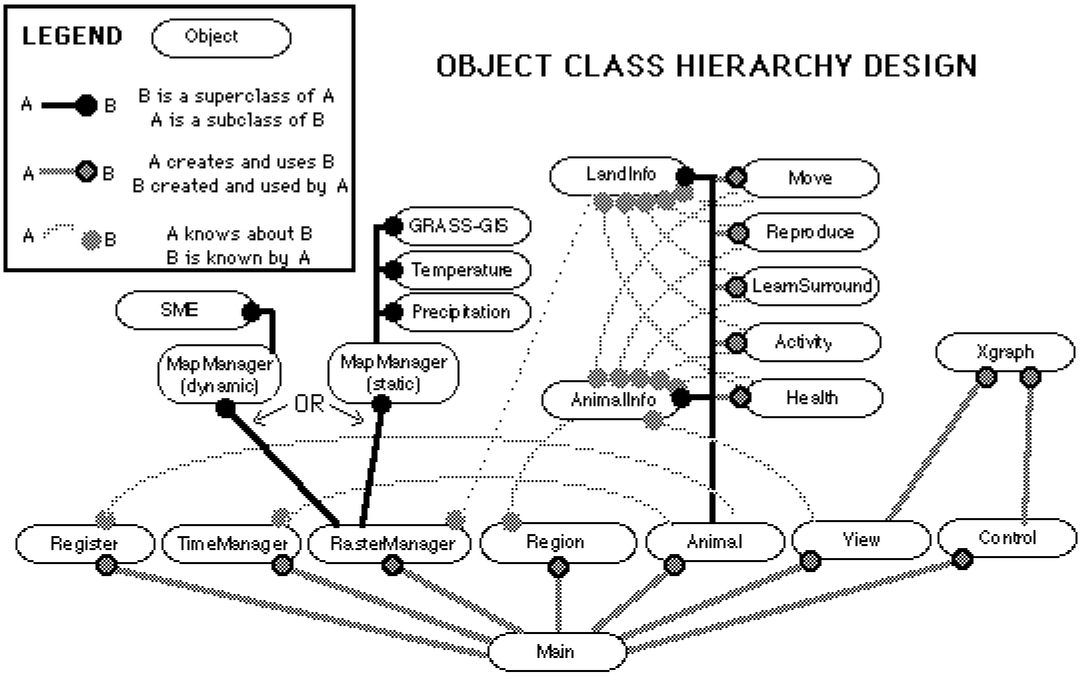


Figure 7: Object Class Hierarchy

Demonstration

A series of proof-of-concept demonstrations has been developed. These demonstrations are inspired by the Desert Tortoise of the Mojave Desert, but must not be viewed as any more than realistic. Their purpose is to show what capabilities are provided by the development concepts discussed in the previous section. Land managers and ecologists may see possibilities for applying these approaches to their particular land management challenges.

Dynamic Entities / Static Land

The purpose of this static example is to demonstrate the ability to generate a dynamic entity that interacts with static GRASS maps accessed directly from native GRASS databases. This demonstration uses GRASS data for a small section of Fort Irwin, California, and the IMPORT/DOME general purpose support classes described above. The animal class for this demonstration was inspired by the Desert Tortoises found on Fort Irwin. This class uses optional support classes, which include health, movement, activity, and "LearnSurround" objects. The key focus of this simulation is the behavior of a number of instantiated animal objects with respect to the weather and the landscape. Males fertilize unfertilized females if their ages, time of year, and state of health is appropriate. Females give birth after a gestation period, to a single young. The simulation begins by opening up the associated GRASS maps and initializing objects that generate the various viewports. It also instantiates support object s including a Region manager, a TimeManager, and an animal Register. Each of the viewports is associated with an internally defined update frequency and these objects schedule their first update. The simulation then instantiates a number (here fifty) of animals. Associated with each animal are support objects instantiated for each individual animal that include movement, health, and activity. Each animal schedules itself for updating.

When an animal updates itself it goes through the following steps. First it checks to see if it is still alive. If not, the animal does nothing and does not schedule itself for any further updates. It then collects information about its surroundings including the other animal entities that fall within it s defined home range radius. This range is represented by a circle surrounding each animal in the MapManager generated viewport (Figures 8 and 9). Other information includes the average slope, elevation, temperature, and precipitation in he home range area. Then, based on temperature, the activity of the animal is determined by the Activity object. If too hot, the animal aestivates and if too cold it hibernates. In either case the animal dynamically sets its update interval to a week; otherwise it is set to a day. It then, through its associated Health object, eats, drinks, and determines changes to its age, weight, length, stress index, and satisfaction index. The Health object maintains information about the current state of the animal's hunger and blood concentration. If the stress level of the animal gets too low it can die based on a death potential that rises with increasing stress. A LearnSurround object then adjusts the animal's satisfaction with the surroundings by allowing it to "learn" a little more about its surroundings. This object simulates the ability for an animal to come to "know" its local terrain over time.

If the animal is of appropriate age, it engages in reproductive activity. For males this means fertilizing females of age who are not already fertilized and exist within his "home range". The probability of fertilizing a female is based on the distance between the female and the center of the male's range and the age of the male. More distant females are less likely to be encountered by the male and therefore less likely to be fertilize d. Males who have just entered their reproductive years are less likely to succeed in fertilizing a female than males in the middle of those years. Similarly, fertilization success decreases as a male gets very old. Fertilized females then give birth after a preset number of days required for gestation. In these animals the baby animals are provided no parental care after birth, as is the case with Desert Tortoises.

Finally the "Move" object then facilitates movement of the animal based on its home range and motivation to move which is a function of stress and satisfaction. After this is all accomplished, the animal schedules itself for another update at a later time in the simulation. The delay time is a function of the animal's activity as determined by the Activity object.

Look now at the two snapshots of the dynamic simulation that result from the above code representing day 3 (Figure 8) and day 155 (Figure 9), a span of about half a year. This demonstration shows that individual entities can be instantiated, placed on the landscape, and then interact with the landscape as defined by static GRASS maps accessed directly out of the GRASS database structure. Some animals have died and are indicated by locations in the second figure that have lost their "home range" circle. Some of the animals have moved very little and others quite a bit, reflecting different satisfactions of the animals with their initial positions. Females have been fertilized and have given birth to animals represented by very small home range circles. On day 155 there is only one fertilized female, number F-15, who is about a quarter of the image up from the bottom and a third of the way across from the left. Her range circle is slightly lighter in grey-scale images and green in colorimages.

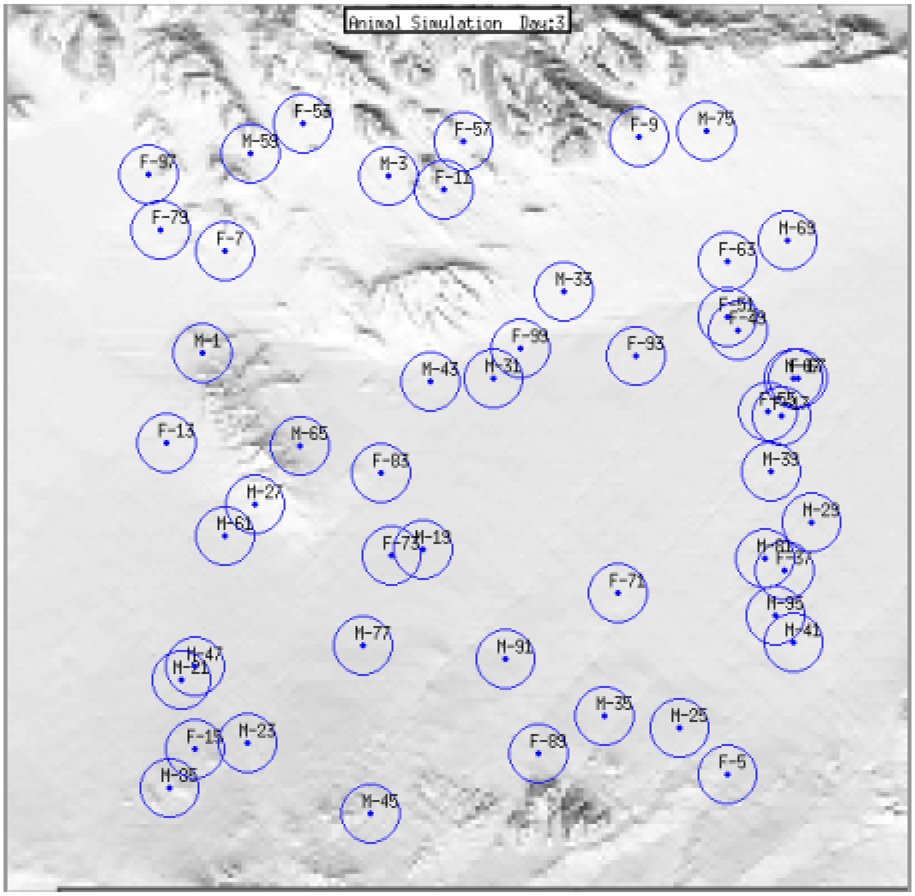


Figure 8: Reproduction: Time Step 003

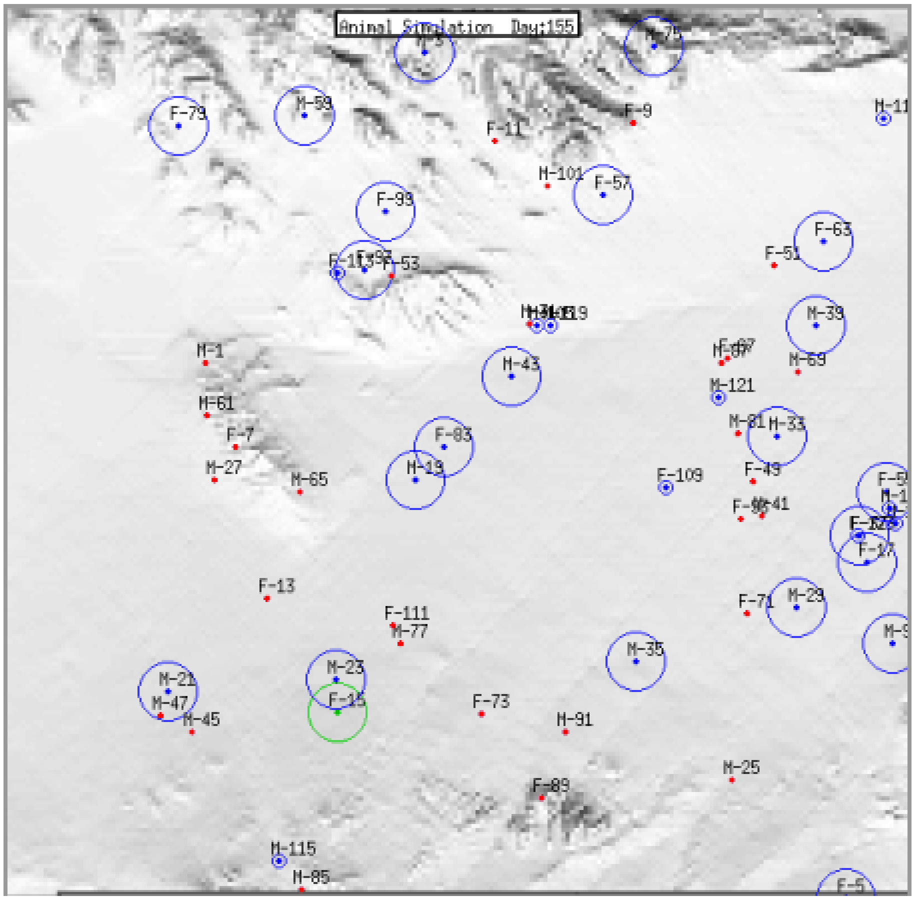


Figure 9: Reproduction: Time Step 155

This simulation can also be run with static landscape maps replaced by a dynamic landscape simulation defined by the Stella model described above (see Figure 3). This model was converted into C++ code using the Spatial Modeling Environment software and then compiled into a single program. IMPORT/DOME controlled the overall timing and facilitated information exchange between the two different simulations. Access to the SME model was facilitated by a dynamic MapManager class.

Predator Prey

The animals of the previous demonstration have been turned into prey animals which have relatively rapid asexual reproduction and short life spans. They still must live off the vegetation represented by the static GRASS green and brown vegetation maps and can die of starvation as well as old age. A new predator animal has also been developed for this demonstration. It is larger, has a bigger range and its hunger, stress, and satisfaction indices. Attacks always result in death to the prey in this demonstration, but that is a decision made by the prey itself. Future demonstrations could easily allow prey to avoid being killed in an attack based on their local terrain, the weather, the predator itself, or other factors peculiar to that animal.

Figures 10 (day 13) and 11 (day 285) are taken from the dynamic landscape simulation based on the predator and prey animals developed. Note for day 13 that the single predator (F-51) has a much larger home range compared to the prey animals. The prey have varying sizes of home ranges based upon their ages, which are initialized at random. One prey animal, F-39, located just south of the predator has been killed by the predator and is associated with a pink dot (very light grey in gray-scale images) indicating a predator death. By day 285 the predator has killed a good number of prey and quite a few other prey have died either of starvation or old age. The prey items have increased in density in some areas of the simulation area, and are completely absent from others. This makes it difficult for the predator to locate prey, but once located it can remain nearby and feed for quite a while. Patterns of prey patches are developing based on a combination of food availability for the prey and predation patterns.

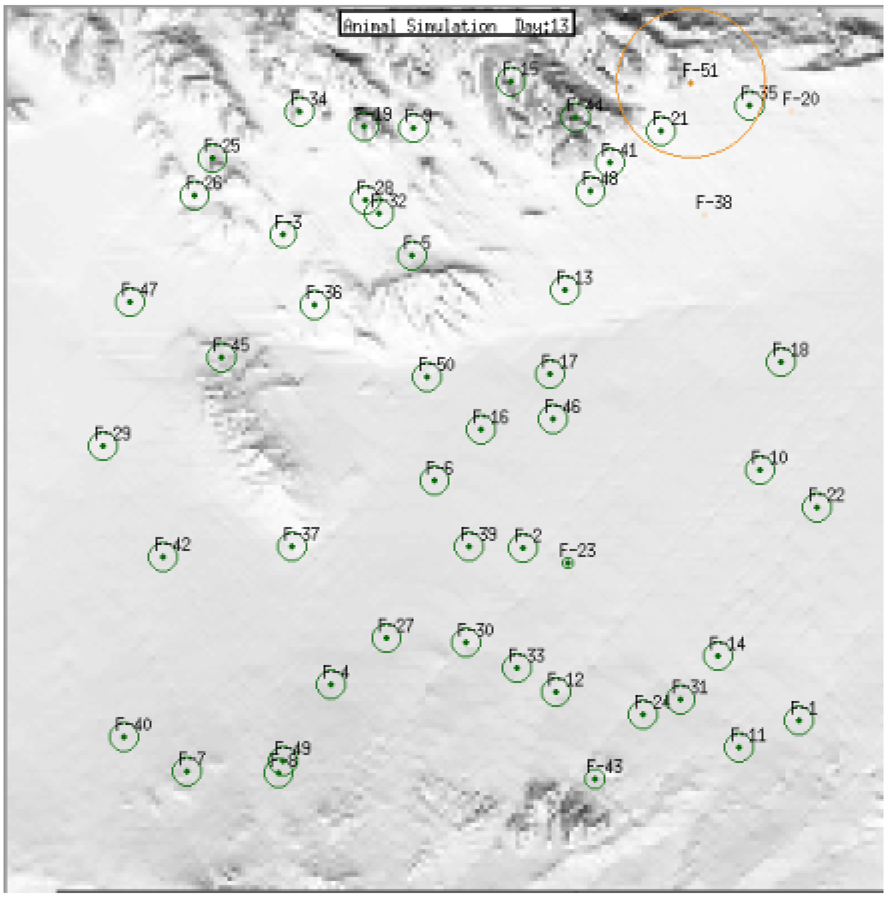


Figure 10: Predator/Prey: Time Step 013

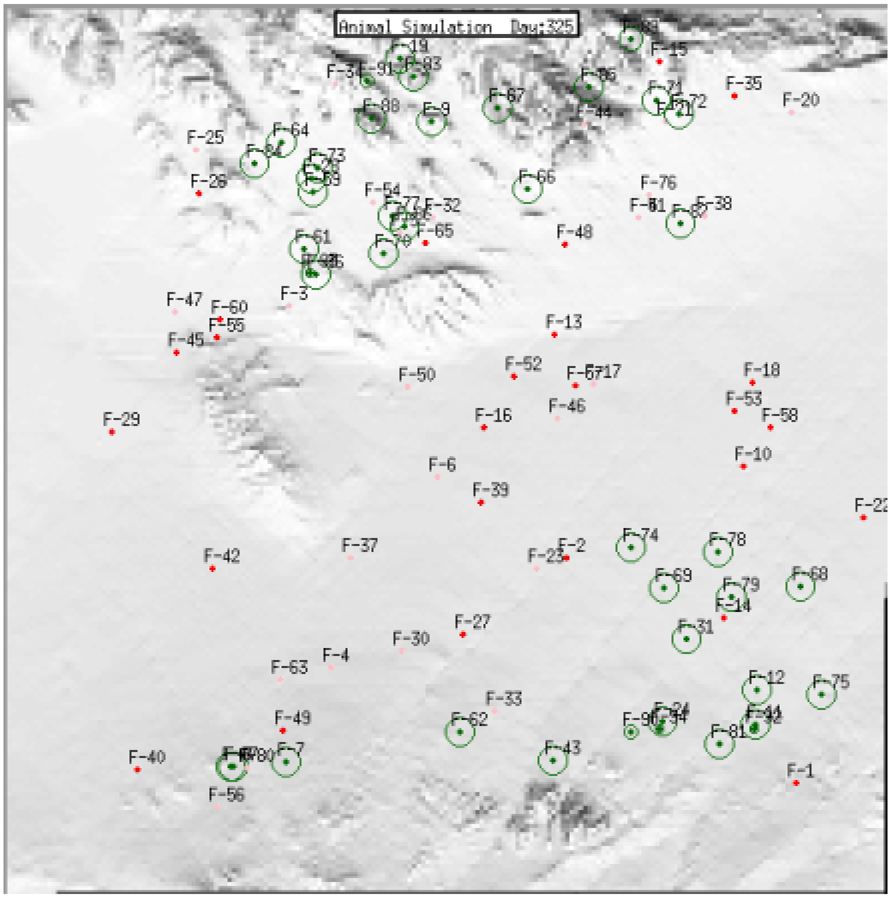


Figure 11: Predator/Prey: Time Step 325

CONCLUSION

This project has developed a fundamental approach to facilitate modeling of dynamic mobile entities in combination with either static or dynamic, raster based landscape simulations. The approach provides significant flexibility in specifying a wide variety of entities captured as classes in the IMPORT/DOME object-oriented dynamic simulation language. At a more general level, an ecological landscape modeling paradigm has been demonstrated that not only allows, but encourages , the capturing of natural processes within simulation environments based on different software. This is the fundamental advantage that must be expected from future geographic modeling systems (GMS). The various simulation components operate separately, but communicate with each other during simulations.

This approach differs from the gap-based forest models wherein individual trees are simulated within square landscape patches, which themselves are arrayed upon the landscape. In those models a neighboring tree is treated differently based on whether or not it falls within the same patch. The square patches are imposed upon the fundamental data structure of the system that captures the location of and relationships between trees. In contrast, the approach to landscape simulation used in this project allows two disparate software simulation environment s to be used, each for what it is best designed to do, yet communicate with each other as appropriate. A powerful raster-based landscape simulation environment was used to capture the dynamics of weather, soil saturation, and vegetation densities, while an equally powerful object-oriented simulation language was used to capture the dynamics associated with individual animals acting within and upon the landscape. Neither environment is suitable for both applications, although a force fit is certainly possible. Neither is captured within the context of the other; each is provided equal status. Future GMS will provide a wide variety of interacting, yet distinct landscape simulation capabilities. Modelers and landscape managers will be able to match landscape processes with the most appropriate simulation module.

Two simulation environments SME, the Spatial Modeling Environment and the new system developed here to support the simulation of the behavior of individual animals were joined. The latter was developed with the IMPORT/DOME dynamic simulation language. Specific capabilities included:

- Dynamic, object specific time steps
- Interaction between animals, including predator-prey and reproduction
- An efficient search mechanism for finding neighbors
- Generic means for evaluating neighbors
- Recognition through evaluation of characteristics
- Ability to add new animals without reprogramming others
- Interaction with static or dynamic landscapes

Nature operates at a variety of different spatial and temporal scales, organized at a number of hierarchical levels. Using the integrated combination of SME and IMPORT/DOME, systems can be developed that explicitly simulate the behavior of organ systems through communities. The simulation focus here is the landscape, a "middle-number" system, which is not analyzed well with statistics, because it lacks a large enough sample size and is not easily simulated as a small set because there are too many distinct parts. The approach developed in this work embraces the hierarchy theory approach to dealing with middle number systems. System components at multiple hierarchy levels can be simultaneously simulated using software approaches that reflect the spatio-temporal scale of each level. No one level dominates the simulation. Instead, each is allowed to communicate with other simulations being conducted at different scales. For example, a simulated individual animal belonging to species A may interact with a population model of species A, an individual model of species B, or a population model of species B.

The power and potential of this combination of simulation approaches opens the door to completely new classes of dynamic, spatial, ecological models. Further, this new simulation environment opens the options for integrating knowledge and results gained through disparate environmental studies. It provides an interdisciplinary forum for formally capturing this information in a manner that increases its potential for affecting land management practices.

REFERENCES

Gaur, N., & Vieux, B. (1992). r.fea (Version GRASS 4.1). Oklahoma City: US Army Construction Engineering Research Laboratory.

Maxwell, T., & Costanza, R. (1993). Spatial Ecosystem Modeling in a Distributed

Computational Environment. In J. van den Berg & J. van der Straaten (Eds.), Concepts, Methods, and Policy for Sustainable Development, : Island Press.

Mitasova, H, Brown, W. M., Gerdesand, D. P., Kosinovsky, I., Baker, T., and Mitas, L. (1994). Modeling spatially and temporally distributed phenomena: New methods and tools for Open GIS, International Journal of GIS (8), 5

Morrison, V. P. (1995). Import/Dome Language Reference Manual (Unofficial software document). ftp pike.cecer.army.mil:/pub/asset/import: Construction Engineering Research Laboratory.

Rewerts, C. C., & Engel, B. A. (1991). ANSWERS on GRASS: Integrating a watershed simulation with a GIS (ASAE Paper 91-2621): ASAE, St. Joseph, MI.

Saghafian, B. (1993). Implementation of a distributed hydrological model within Geographical Resources Analysis Support System (GRASS). Paper presented at the Second International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modeling, Breckenridge, Colorado.

Vieux, B., & Westervelt, J. (1992). Finite element modeling of storm water funoff using GRASS GIS. Paper presented at the Computing in Civil Engineering and Geographic Information Systems Symposium, Dallas, Texas.

Westervelt, J. D., Hannon, B., Levi, S., & Harper, S. (in press). A Landscape Simulation Demonstration of the Desert Tortoise Habitat at Fort Irwin, California (USACERL Technical Report): US Army Corps of Engineers' Construction Engineering Research Laboratories.

Westervelt, J. D., Shapiro, M., Goran, W. D., & Gerdes, D. P. (1992). Geographic Resources Analysis Support System (GRASS) Version 4.0 User's Reference Manual : US Army Corps of Engineers Construction Engineering Research Laboratory.

James D. Westervelt
PI, Construction Engineering Research Laboratories
Champaign, Illinois 61824
j-westervelt@cecer.army.mil

Lewis D. Hopkins
Department Head, Urban and Regional Planning
111 Temple Buell Hall
Champaign, Illinois 61820
ldhopkin@uiuc.edu

Joseph Fall and Andrew Fall

SELES: A Spatially Explicit Landscape Event Simulator

Landscape structure results from complex interactions between geomorphology, climate, disturbance and succession. SELES is a tool for building spatially explicit simulations to model the role of disturbance in creating and maintaining landscape structure. Models built with SELES are raster-based, semi-Markov, whole landscape models (sensu Baker 1989) which use probabilistic disturbance spread. Each raster cell represents a sub-model, with landscape patterns emerging from the accumulation of cell behaviors. The modeler defines landscape events (such as disturbance and succession), which determine the landscape dynamics. Each event is characterized by a set of expressions that govern how often the event occurs, how the cell state is affected by the event, and how the event spreads from one cell to another. These expressions are dependent on one or more cell state variables (vegetation cover, soil type, elevation, slope, aspect, etc.). SELES models can incorporate these state raster layers from a variety of sources and uses PAMAP GIS directly as a spatial database.

Starting with a simple model that operates on a neutral landscape (sensu Gardner *et al.* 1987), the modeler can add levels of complexity incrementally and incorporate parameters from real landscapes to refine hypotheses about landscape dynamics. We describe the event-based simulation engine and how to use it to specify landscape events. We then show how SELES is applied to model the consequences of fire suppression on forested landscape dynamics.

---

1 Introduction

SELES is a tool for building spatially explicit landscape simulation models. These models are useful for exploring the effect of natural disturbance and succession on landscape structure. The purpose of a model built with SELES is not to predict when a particular site will be disturbed or how large or intense a particular disturbance will be. Rather, SELES models are intended for developing and testing hypotheses about landscape dynamics and characterizing the spatial patterns that emerge from a given disturbance regime.

SELES is not tailored for any particular scale, type of landscape, or set of processes, thus it is useful for a variety of landscape modeling projects. The models may be simple and largely deterministic or very complex and stochastic. Each model requires detailed parameterization for modeling different landscapes or scenarios. The user interface provides a straight forward method of specifying the parameters and input data, however, the modeler must carefully choose the scale of the model to be appropriate for the landscape being investigated.

In this paper, we describe how SELES models are built and simulated. We first provide the background motivation that inspired us to develop SELES. A topdown system overview is followed by a more detailed operational description. The final two sections describe an example simulation and SELES' integration with GIS.

2 Background and Motivation

Large-scale vegetation patterns are created by the complex interactions between geomorphology, climate, disturbance, and succession (Forman and Gordon, 1986). The spatial distribution of a landscape feature (e.g. a particular species mix, habitat type, or age class) is dependent on some or all of these ecological effects. The relationship between landscape patterns and the processes that create them is one of the focal issues in landscape ecology (Turner, 1989). Past research efforts have largely focused on developing methods and indices for measuring, describing and differentiating landscape patterns (see Turner, 1989 for review). More recent studies have focused on hypothesis testing and have uncovered startling implications for landscape planning (e.g. Hall *et al.*, 1991; Mladenoff *et al.*, 1994; Spies *et al.*, 1994).

The difference between real landscape patterns and the predictions of a model are one measure of the model's ability to predict landscape structure (Gardner *et al.*, 1987). Turner *et al.* (1989) used a simple model to simulate disturbances on neutral landscapes. The disturbances are modeled as random events that occur with a given frequency (probability of initiating) and intensity (probability of spreading to neighboring cells) (Turner *et al.*, 1989). This disturbance model provides a baseline that can be used to measure the improvement in predictability that is achieved by modeling geomorphologic, climatic, and biotic effects. Our objective is to develop a tool that will allow different effects (such as topography, soil, or climate) to be incrementally added to the model. These extensions will provide a means of testing hypotheses that account for the differences between the observed and expected patterns.

3 System Overview

Landscape structures in SELES (vegetation coverage, topography, soils, etc.) are represented by rasters (a grid of fixed size square cells). PAMAP GIS is used as a spatial database to supply these raster layers. Each cell in the landscape is characterized solely by the structural attributes (values of raster layers) at its location and is assumed to be spatially and compositionally homogeneous. Patches and spatial relationships are implicitly expressed by the arrangement of cells with similar characteristics. Thus, landscape pattern is not modeled directly, but is an emergent property of the cell level processes.

All landscape processes, such as disturbance and succession, are defined by the model builder as *landscape events*. Landscape events are the driving forces of landscape dynamics in SELES models and provide the mechanism for changing landscape structure over time. Models are constructed in SELES by defining one or more landscape events and providing an initial state (set of raster layers representing initial landscape structure). The initiation and spread of a disturbance events is probabilistic. The probability of disturbance in any particular cell is computed dynamically, based on the current state of the cell. The simulation models the effect of these landscape processes on landscape structure over time. Intermediate results may be analyzed to determine how structure changes over time and Monte-Carlo simulations may be used to determine how much variation in structure might be expected from the same set of landscape processes.

3.1 Abstract Simulation

The following is an abstract description of the steps required to define and execute a SELES simulation:

1. Define a landscape. This will require the modeler to select one or more raster layers to be used in the model. One such layer, the landscape features, is always required - it contains the "living" cover acted on by the model. These layers are then employed to describe a landscape feature's susceptibility to landscape events (usually the probability of being disturbed).

2. Define one or more landscape events. Most SELES models will have at least two landscape event types - one for succession and one to characterize the disturbance regime. Each landscape event is defined by its return interval, frequency, intensity, and spreading characteristics. The modeler must also specify a transition table for each event, which determines how the event will cause the landscape features to change state.

3. The simulation is executed for a specified time interval. During the simulation the following steps take place:
    1. Each landscape event (e.g. disturbance) is scheduled to occur at some future time, based on its return interval.

    2. When a landscape event occurs, a probability of initiating is computed for each cell, based on the cell's current state. Cell events are randomly initiated in a number a cells, based on this probability.

    3. A cell event may cause the cell to change states (make a transition from one landscape feature to another), if the event was intense enough.

    4. The cell event may also spread to each of the four neighboring cells, depending on its spread characteristics.

    5. Subsequent events will operate on the modified landscape features.

The above abstract simulation is depicted in a conceptual diagram (Figure 1), which shows an initial landscape, a fire caused by a lightning strike and the resulting landscape after the disturbance event and one succession event.

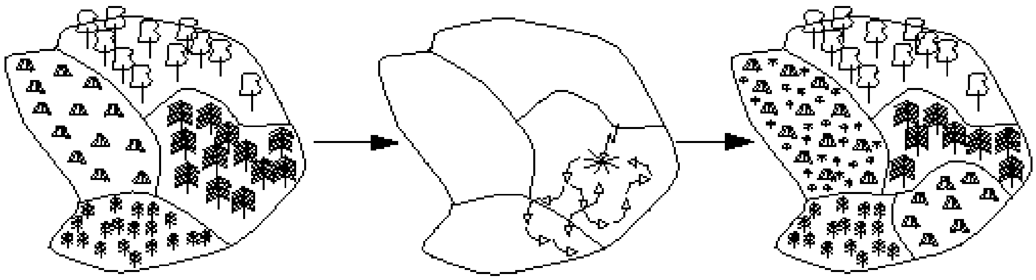


Fig. 1: Conceptual diagram of fire initiation and spread

4 System Details

The following sections describe the model components in more detail.

4.1 Cell Sub-Model

Each cell may be thought of as a sub-model (Figure 2). This cell-model defines the behaviour of a single cell. A relatively simple cell-model was chosen to facilitate parameterization, comprehension, and analysis of the model output. Each cell contains only two pieces of variable state information:

1. The landscape feature that typifies the cell (e.g. dominant vegetation cover).

2. The time at which the landscape feature last changed states, which can be used to compute an age or time since transition for the cell.

In addition to this variable state, each cell is further characterized by zero or more static state parameters, such as the cell's elevation, soil type, or precipitation.

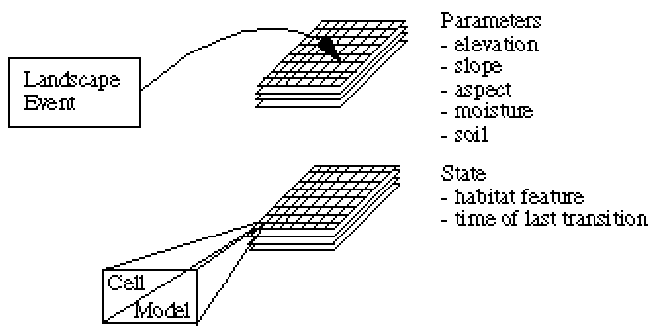


Fig. 2: Conceptual diagram of the relationship between landscape events and cell sub-models

4.2 Landscape Events

Landscape events describe the forcing function(s) of SELES models. Both disturbance and succession can be thought of as events that occur on the landscape. Although these events have different effects, both may be described with a similar set of parameters. In a SELES model, any operation that changes the state of a cell is defined as a landscape event. A landscape event is characterized by six properties:

1. *Return time.* This is a probability distribution that specifies the time interval between events of this type. The event handler uses a random return time from this distribution to re-schedule the event.

2. *Probability of initiating.* This is a formula that defines the probability that the event will affect a given cell. The formula is a function of the cell model state.

3. *Initial intensity.* This is a probability distribution that specifies how intense the event will be. When a disturbance has initiated in a cell, it is given a random initial intensity from this distribution.

4. *Susceptibility*. This is a formula that defines how susceptible a cell is to the event. This formula is a function of the cell model state. This susceptibility is combined with the initial intensity to compute a final intensity for the event in each cell.

5. *Spread characteristics.* If an event is a spreading event, then there is a probability distribution that defines the spread time-step. This refers to the length of time required for the event to propagate across a cell and spread to its neighboring cells.

6. *Transition Tables.* State transitions (Lertzman, 1994; Moore & Noble, 1993) in SELES are implemented as semi-Markovian transition tables (Baker, 1989). Associated with the transition table is a threshold table. Element j in the threshold table contains a minimum event intensity that must be achieved before a cell in state j may make a transition. Note

that a SELES model may run as a pure Markovian model if all transition thresholds are set to 0. Transition matrices of this nature may be derived using the method of Hall *et al.* (1991).

4.3 Landscape Event Simulation

The simulation is controlled by a discrete-event simulation engine (DES). The DES uses a priority queue to maintain all future events in their chronological order. It runs asynchronously - the current time is always the time of the event at the head of the queue. Events are stored on the queue as with a handling routine and a single parameter. The DES simply removes the next event from the head of the queue, calls the handling routine for the event with its parameter, waits for the handler to finish, then removes the next event. New events are generally added to the queue by the event handlers themselves.

4.3.1 Handling Landscape Events

During the simulation, instances of landscape events are processed by the *landscape event handler*. Its parameter is simply the landscape event description (section 4.2). The landscape event handler performs two tasks:

1. It computes a *probability of initiating* for every cell. It uses a stochastic process based of this probability to determine if the event should be initiated in a cell. If so, the event is initiated by scheduling a *cell event*.

2. It schedules the next landscape event according to the *return time*.

4.3.2 Handling Cell Events

Cell events are initiated either by landscape events (as described above) or by spreading from neighboring cells. The *cell event handler* performs two tasks:

1. Using the event's initial intensity, the cell's susceptibility, and the transition table, it determines if the cell should change state. If so, the landscape feature layer is updated.

2. If the event is a spreading event, then cell events are scheduled in each of the four cardinal neighboring cells.

4.4 Assumptions

SELES was developed with the following assumptions and simplifications:

1. Disturbance can only spread to adjacent cells (i.e. no "jumping" over cells). However, a disturbance may pass through a cell without affecting it.

2. Disturbances will spread to cardinal neighbors only (i.e. no diagonal spread).

3. A spreading disturbance will not spread back to its originating cell.

4. The rate of spread is not dependent on the cell state.
5. All data layers (rasters) have the same grain and extent.
6. Seasonal and annual fluctuations are not considered explicitly.
7. There are a finite number of discrete states that may occupy the feature layer.
8. State transitions can be defined by a simple transition vector that is independent of all cell state other than the cell's current feature. In reality, transitions may depend on other characteristics of a cell (e.g. elevation may have an effect on succession transitions).
9. State changes will only affect the feature layer. All other landscape structures (e.g. topography, soil, climate) are static for the duration of the simulation.
10. Landscape processes can be defined by a simple set of stochastic equations describing their behaviour (frequency and intensity) for all possible landscape states. Their spatial and temporal extents are emergent properties.

5 Example: Fire Suppression in a Western Hemlock/Douglas Fir Landscape

This example is intended to demonstrate the use of SELES in a management scenario. Suppose we are designing a fire suppression policy in a Western Hemlock/Douglas Fir forested landscape. We would like to compare possible results of fire suppression on landscape structure with a base-case scenario (no fire suppression). The model we developed is hypothetical - it is not based on any particular location. It is intended to aid in the development of hypotheses about the potential consequences of fire suppression on forest landscape structure.

For our model, we assume that forest cover can be divided into five feature types: disturbed, young Douglas Fir (yDF), mature Douglas Fir (mDF), mixed mature Douglas Fir/Western Hemlock (mWH/DF), and mature Western Hemlock (mWH). Succession is modeled as an annual event (*return time* of 365.25 days). The event will simply increase the time since last transition, and then attempt to make a succession transition. Figure 3 shows how we describe succession in this model. The year on the transition arc indicates how long a cell must remain in the source state before succeeding. The percentage specifies the probability of transition once a site reaches the minimum age. By specifying 100% for all transitions, succession will be purely deterministic. This behaviour can be specified by setting up the succession landscape event appropriately.

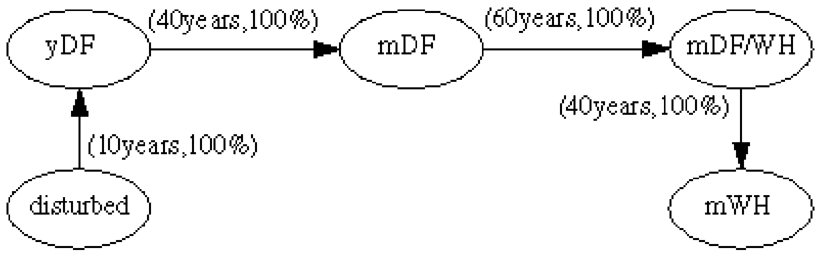


Fig 3: Succession Transitions

The disturbance regime for the model consists of lightning caused fires with the following characteristics:

1. The *return time* is normally distributed about 1000 days with a standard deviation of 100 days.
2. All sites have the same *probability of initiating* during a disturbance event (0.0005).
3. The *initiating intensity* of a lightning strike is normally distributed about 0.8 with a standard deviation of 0.1.
4. The *susceptibility* of features increases with successional stature: 0 for the disturbed feature, 0.3 for young DF, 0.4 for mature DF, 0.45 for mixed DF/WH and 0.49 for mature WH.
5. Fires will *spread* to neighboring cells after 1 day.

The transitions for disturbance are illustrated in Figure 4. We assume that mature DF is tolerant to the type of fires in this landscape, while WH is intolerant. A site will only change state if the final intensity is more than 0.5 (except for the disturbed feature, which is not affected by fire). In case the final intensity is larger than 0.5, cells of type young DF and mature WH will change to the disturbed feature, while cells of type mature DF and mixed DF/WH will be reset to mature DF. These transitions are also all deterministic.

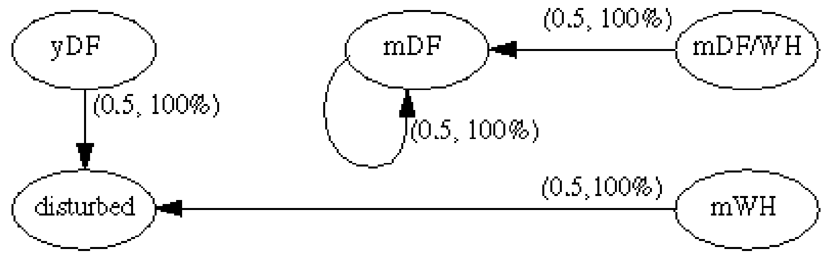
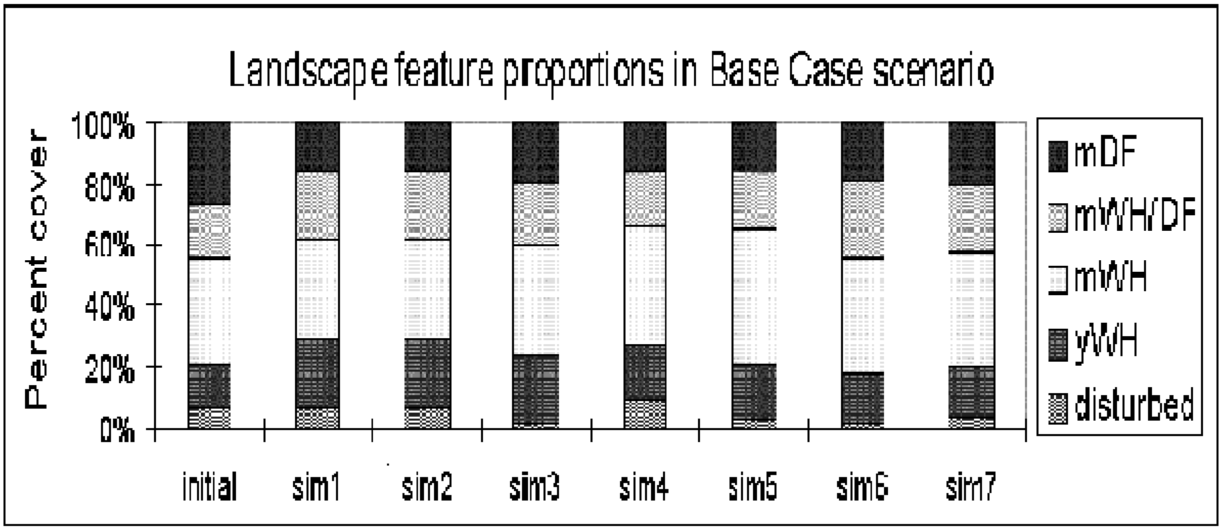


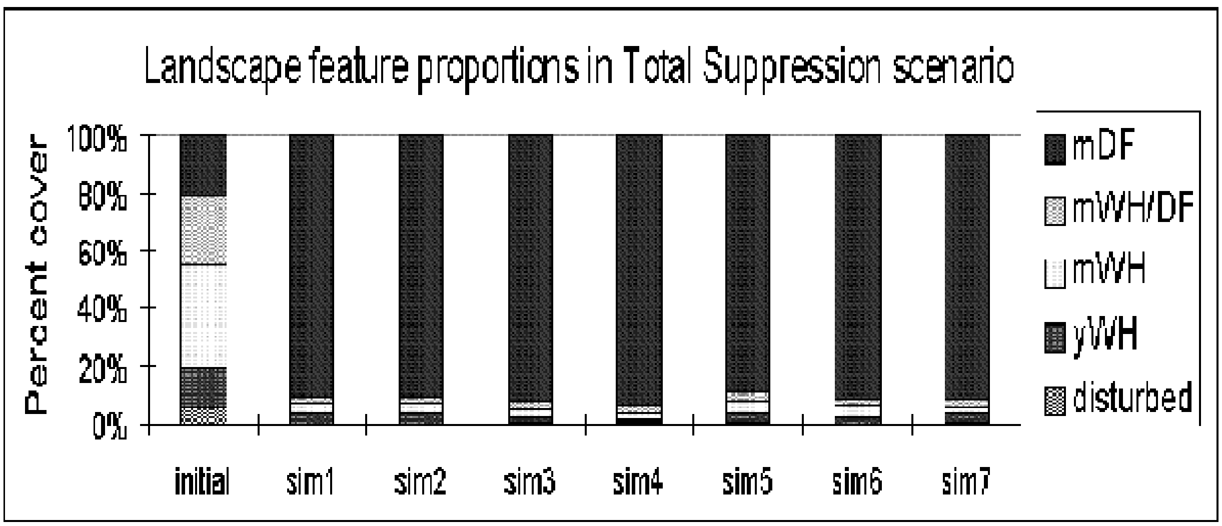
Fig 4: Disturbance Transitions

The above scenario describes the natural disturbance and succession regime. Our management goal is to explore the potential impact of a fire suppression policy. To this end, two other scenarios are developed to model possible results of the policy. In the first alternative model we will assume that the suppression policy is completely effective in stopping or reducing the extent of all fires that are started. To model this we must change the disturbance regime. To this end, we reduce both the *probability of initiating* (to 0.0002) and the *initial intensity* (to a mean of 0.54 and a standard deviation of 0.01) in the disturbance event definition. Reducing the probability that a cell will be disturbed accounts for fires that are extinguished before they really get started (i.e. before they affect a cell's state). Reducing the disturbance intensity account for the fire fighting activities mounted against larger fires, by decreasing their potential to spread.

As a second alternative model we will assume that although the suppression policy may be effective, there may still be fires that occasionally get out of control. In this scenario, we use the original fire regime, but increase the *return time* by a factor of 20 to model a premise that one out of 20 fire events will get out of control. We assume that the fires which are successfully suppressed do not contribute significantly to landscape structure, thus we do not model these small fires at all.

We call these three scenarios the Base Case, Total Suppression and Partial Suppression scenarios, respectively. We ran each scenario 7 times for 300 years per run on the same initial landscape. The relative proportions of features on the initial landscape and the final landscapes for each of the 7 runs is graphed in Figure 5.


Landscape feature proportions in Base Case scenario
Percent cover
100%
80%
60%
40%
20%
0%
initial
sim1
sim2
sim3
sim4
sim5
sim6
sim7
mDF
mWH/DF
mWH
yWH
disturbed


Landscape feature proportions in Total Suppression scenario
Percent cover
100%
80%
60%
40%
20%
0%
initial
sim1
sim2
sim3
sim4
sim5
sim6
sim7
mDF
mWH/DF
mWH
yWH
disturbed

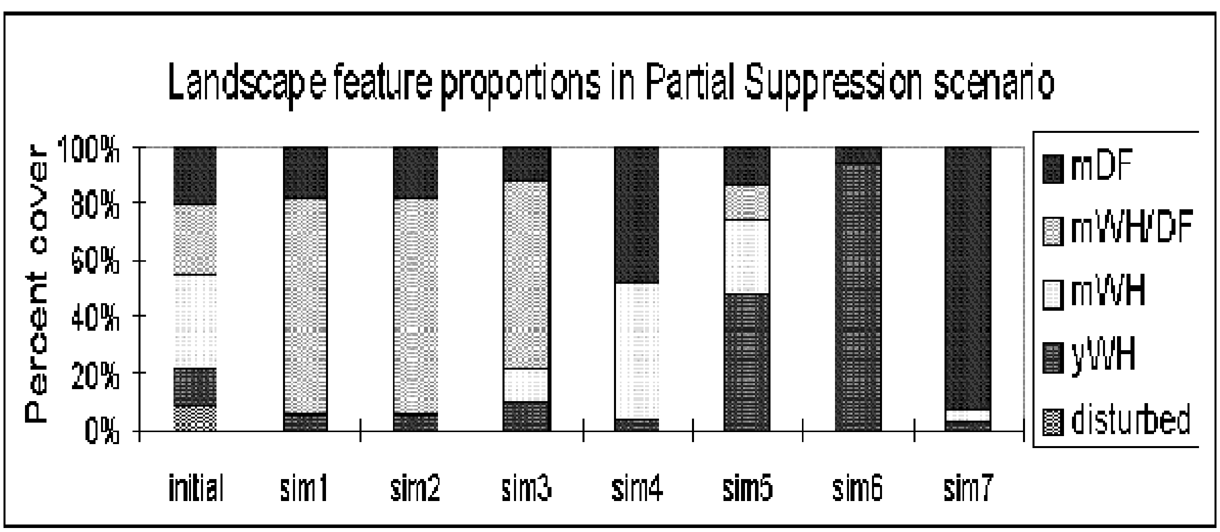

Landscape feature proportions in Partial Suppression scenario
Percent cover
100%
80%
60%
40%
20%
0%
initial
sim1
sim2
sim3
sim4
sim5
sim6
sim7
mDF
mWH/DF
mWH
yWH
disturbed

Fig 5: Relative proportions of features on the initial landscape and 7 final landscapes after 300 year simulation. Top graph (a) is for the Base Case scenario. The middle (b) and lower (c) graphs are for the two fire suppression scenarios (see text).

In the base case, the relative proportions of features is approximately the same in the initial landscape and all final landscapes. Figure 6 shows the initial and one of the final landscapes for this scenario. Visually analysis show that these landscapes ``look" similar, perhaps with slightly different average patch sizes.

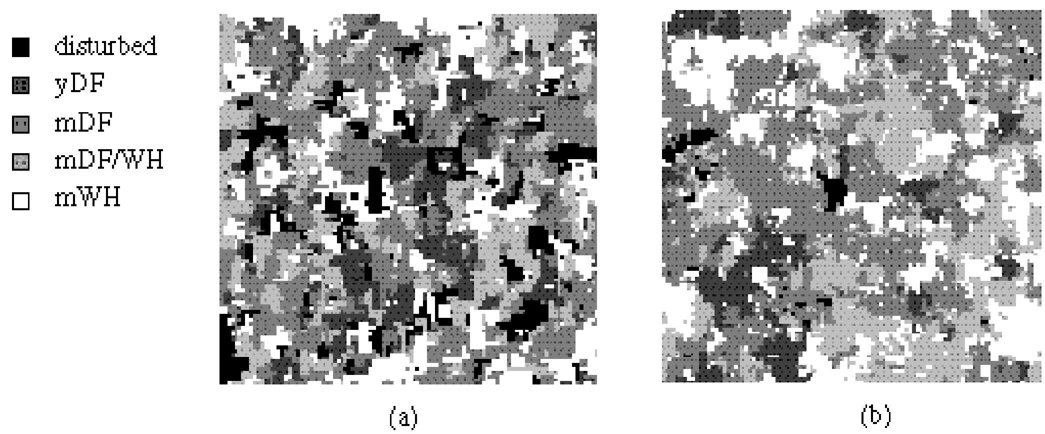


Fig 6: Natural Disturbance landscape. Initial state (a) & Output state after 300 years (b)

The total suppression policy, however, removes large fires as a major component of landscape dynamics, allowing most of the landscape to succeed into mature WH. Due to the less frequent, less extensive nature of fires in this case, the only difference from a scenario with no fires is that about 10% of the landscape is in small patches at various stages of development. Figure 7 shows a sample landscape after 300 years starting with the same initial state as in Figure 6.

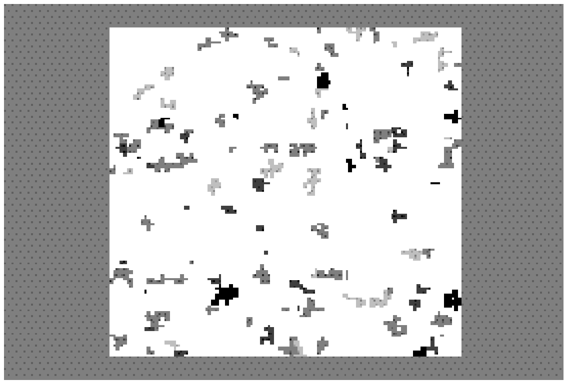


Fig 7: Total fire suppression scenario after 300 years

The partial suppression scenario contrasts with the total fire suppression scenario. As can be seen in the graphs (Figure 5), total suppression causes highly variable landscape structure. This instability is caused by fire suppression, which allows the forest to succeed into mature WH between fire years. When a fire does get out of control, the higher connectivity of WH (the most susceptible feature) causes large areas to burn and reset to the disturbed feature. Thus, over time the landscape changes radically and feature proportions are highly variable. Figure 8 shows three snapshots taken during one simulation, each of which contains substantially different proportions of features (starting from the initial landscape in Figure 6). These landscapes do, however, have in common a greater simplicity of structure than in the base case.

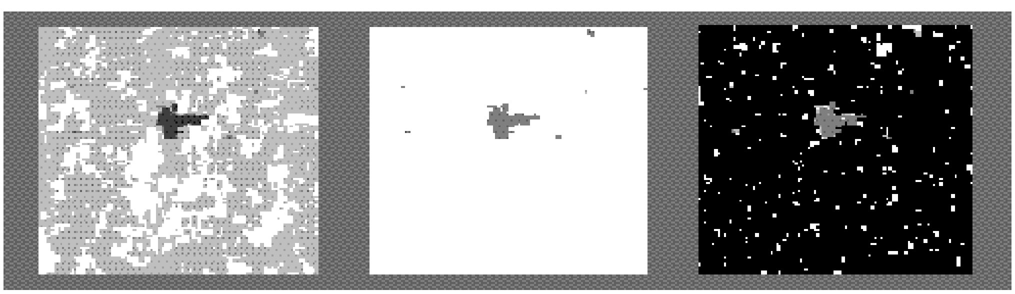


Fig 8: Total fire suppression scenario after 260, 280 and 300 years

What can we learn from these models? Natural disturbance regimes may very well maintain critical landscape structures (Figures 5(a) and 6). By intervening with the role of natural disturbance, we may dramatically alter landscape patterns (Figures 5(b) and 7). However, our management actions may also have a de-stabilizing effect (Figures 5(c) and 8). The interactions between disturbance, succession and landscape pattern are very complex. In scenario 1, the spread of fires was controlled by a patchy landscape structure. In scenario 2, a change in disturbance regime caused the landscape to become much more heterogeneous, with a few small patches occurring in a Western Hemlock matrix. In scenario 3, the homogeneity of the landscape allows disturbance to propagate much more easily. Thus, a large portion of the landscape tends to be affected when a disturbance does occur.

6 GIS Integration

The modelling environment for SELES requires the following characteristics:

1. A spatial database to provide input layers and store results.
2. A general purpose discrete-event simulation engine.
3. A visualization tool for visual representation and analysis.
4. A spatial statistics package for quantitative analysis of results.
5. A customizable user interface for model parameterization.
6. A programming language for custom event-handling modules and model integration.

Unfortunately, it is not possible to find a single package with all these features. PAMAP GIS recently integrated their product with a scientific visualization tool (AVS). This package also provides a modular programming environment, using the C programming language, and a simple configurable user-interface. We used PAMAP/AVS to implement SELES, however, all modelling tools had to be custom built. Quantitative analysis of results is done through a third party landscape statistics package, FRAGSTAT (McGarigal & Marks, 1993), which was converted to operate in our modeling environment. To ease simulation model development using GIS, it would be very useful to have a general purpose, discrete-event simulation engine incorporated as an integral GIS tool. The spatial statistics package should also be available within the GIS environment to facilitate quantitative analysis.

Besides the lack of an integrated simulation modelling environment, there are several issues that would help make SELES more useful:

1. While model results may be stored back to the spatial database, the GIS currently treats the model results the same as any other layer in the data set. While this is not a problem per se, it would be useful if modelling results might be segregated from the rest of the data set, while still maintaining the spatial relationships with the original data.
2. GIS does not currently handle time-series datasets well. It would be useful to store intermediate model results so that the simulation could be played back, but the volume of data makes this impractical. However, between time steps there is often very little difference in the state raster. The ability to store a time-series dataset as an initial state plus a set of differences for each time-step would be very useful for analyzing simulations.

3. We would like to experiment with the PAMAP GIS dynamic buffering program (SPREAD) as an alternative to our custom spreading algorithm, but have found that programmatic access to GIS algorithms is difficult. We have not explored how GIS operations might help us analyze model results.

4. It would be nice to model landscape structure on multiple scales. We need multi-resolution data structures: rasters that allow more detail in some cells or layers than others.

5. Polygons boundaries are currently very "hard", whereas hard boundaries are rarely found in natural landscapes. Continued work on "fuzzy" polygons will be important.

6. Polygon definitions are equally "hard". We need "fuzzy" polygon definition so that polygons that are in similar, but not identical, states can be treated as either different or identical.

In the end, GIS and modelling should impact the policy and decision making processes. Modelling results (such as the example above) will be of use to policy makers only if the following conditions are met:

1. The model must be straightforward and simple to understand. In this respect SELES fails. It is intended for use by scientists, not policy makers.

2. All the modelling tools must be seamlessly integrated so that it can be presented as a single tool.

3. Data formats must be standardized so that data can be used from multiple sources without the attendant data transfer nightmares we are all familiar with.

4. The model must have a simple, intuitive user interface to make it accessible.

7 Conclusion

There are tradeoffs between the generality of a tool and how well it performs in specialized situations. We have developed a general landscape simulation tool while attempting to maintain its utility. As such there are areas in which SELES could be improved to handle particular simulation scenarios. Below are several important unresolved issues. Some of these can be viewed as areas for future research, while others are inherent in the design of SELES.

1. SELES models are potentially complex to set up. An understanding of the interactions among the components of landscape events and transitions matrices is essential to generate the desired behavior.

2. There are some model parameters for which it may be difficult to establish values based on empirical data. These include susceptibility, initiation intensity and transition thresholds. These three parameters are also closely related, further compounding this problem.

3. Running SELES simulations is computationally expensive: long runs of Monte-Carlo simulations may be infeasible (a 300 year simulation took approximately 10 minutes on a dedicated Sparc 2 workstation)

4. It is difficult to incorporate planned (purely deterministic) events into the model (e.g. prescribed burns, logging).

However, it appears the tool is useful for testing more complex hypotheses than the simpler models that motivated us. Future work should concentrate on establishing techniques for deriving SELES parameters from empirical data. These techniques will allow us to develop models for specific landscapes.

Acknowledgements

We thank Ken Lertman for encouraging us to explore the potential of SELES, and Brigitte Dorner for many helpful discussions on both the development of SELES and the production of this paper. This research was supported by Andrew Fall's ECO-Research Doctoral Fellowship, Joseph Fall's NSERC PGS-A Postgraduate Scholarship. Additional support was provided by Veronica Dahl's NSERC Research Grant 31-611024 and NSERC Infrastructure and Equipment Grant given to the Logic and Functional Programming Lab. We are also thankful for the use of facilities provided to us by the Forest Ecology Lab in the School of Resource and Environmental Management, and by the School of Computing Science at Simon Fraser University.

References

Baker, W. L. A Review of Models of Landscape Change. Landscape Ecology, 2(2), 111-133. 1989.

Forman, R. T. T., & Gordon, M. Landscape Ecology. New York: John Wiley and Sons. 1986

Gardner, R. H., Milne, B. T., Turner, M. G., & O'Neill, R. V. Neutral Models for the Analysis of Broad-Scale Landscape Pattern. Landscape Ecology, 1(1), 19-28. 1987.

Hall, F. G., Botkin, D. B., Strebel, D. E., Woods, K. D., & Goetz, S. J. Large-scale Patterns of Forest Succession as Determined by Remote Sensing. Ecology, 72(2), 628-640. 1991.

Lertzman, K. P. Forest Dynamics, Differential Mortality and Variable Recruitment Probabilities. Journal of Vegetation Science, 5. 1994.

McGarigal, K., & Marks, B. Fragstat: A Spatial Pattern Analysis Program for Quantifying Landscape Structure. Unpublished software. 1993.

Mladenoff, D. J., White, M. A., Crow, T. R., & Pastor, J. P. Applying Principles of Landscape Design and Management to Integrate Old-Growth Forest Enhancement and Commodity Use.

Conservation Biology, 8(3), 752-762. 1994.

Moore, A. D., & Noble, I. R. Automatic Model Simplification: the Generation of Replacement Sequences and Their Use in Vegetation Modelling. Ecological Modeling, 70, 137-157. 1993.

Spies, T. A., Ripple, W. J., & Bradshaw, G. A. Dynamics and Patterns of a Managed Coniferous Forest Landscape in Oregon. Ecological Applications, 4(3), 555-568. 1994.

Turner, M. G. Landscape Ecology: the Effect of Pattern on Process. Annual Review of Ecological Systems, 20, 171-197. 1989.

Turner, M. G., Gardner, R. H., Dale, V. H., & O'Neill, R. V. Predicting the Spread of Disturbance Across Heterogeneous Landscapes. OIKOS, 55, 121-129. 1989.

Andrew Fall
School of Computing Science
Simon Fraser University
Burnaby, B.C. V5A-1S6
e-mail: fall@cs.sfu.ca

Joseph Fall
School of Resource and Environmental Management
Simon Fraser University
Burnaby, B.C. V5A-1S6
e-mail: jfall@sfu.ca

Keith C. Clarke, Stacy Hoppen, and Leonard J. Gaydos

Methods And Techniques for Rigorous Calibration of a Cellular Automaton Model of Urban Growth

Abstract

Several lessons about the process of calibration were learned during development of a self-modifying cellular automaton model to predict urban growth. This model, part of a global change research project on human-induced land transformations, was used to predict the spatial extent of urban growth 100 years into the future. The context of the prediction was to evaluate urban environmental disturbances such as land use conversion, urban heat island intensification, and greenhouse gas generation. Using data for the San Francisco Bay area as a test case, methods were developed, including interactive and statistical versions of the model, animation and visualization tools, automated testing methods, and Monte Carlo simulations. This presentation will enumerate, analyze, and discuss the lessons learned during the extensive process of model calibration. Experience with the methods developed may have broader use in assisting the rigorous calibration for other CA models, and perhaps those coupled environmental models with an extensive spatial data component. These methods are now under test as the project moves to a new data set for the Washington, D.C.-Baltimore area.

Introduction

A cellular automaton urban growth model was developed and calibrated as part of the USGS Human-Induced Land Transformations project (HILT), part of a contribution to the U.S. Global Change Research Program (USGS, 1994). The study seeks to understand the urban transition from an historical and a multi-scale perspective sufficient to model and predict regional patterns of urbanization 100 years into the future (Kirtland et al., 1994).

The model provides regional predictions of urban extent to be used as a basis for assessment of the ecological and climatic impacts of urban change and estimation of the sustainable level of urbanization in a region. This paper examines the initial calibration of the model using data from the San Francisco Bay area, and analyzes the approach taken and difficulties encountered in a broad modeling context. Future plans include testing the transferability of the model through calibration in other regions, initally a second data set for the Washington DC/Baltimore area, and extending the model to other scales and a broader set of land cover changes.

The Urban Transition Model

The urban model we developed is a scale-independent, cellular automaton model (Couclelis, 1985; Batty and Xie, 1994a; 1994b, White and Engelen, 1992). The growth rules are uniform throughout a gridded representation of geographic space and are applied on a cell-by-cell basis. All cells in the array are synchronously updated at the end of each time period. The basic principles of the growth rules, based on the Clarke wildfire model (Clarke et al., 1995), are general enough to describe any kind of organic expansion; they have been modified in this model to describe urban expansion. The growth rules are integral to the data set being used because they are defined in terms of the physical nature of the location under study, thus producing a scale-independent model.

The model was implemented as a computer program written in the C programming language. It operates as a set of nested loops: the outer control loop repeatedly executes each growth "history," retaining cumulative statistical data, while the inner loop executes the growth rules for a single "year." The starting point for urban growth is an input layer of "seed" cells, the urban extent for a particular year identified from historical maps, atlases, and other sources. The growth rules are applied on a cell-by-cell basis and the array is synchronously updated. The modified array forms the basis for urban expansion in each succeeding year. Potential cells for urbanization are selected at random and the growth rules evaluate the properties of the cell and its neighbors (e.g., whether or not they are already urban, what their topographic slope is, how close they are to a road). The decision to urbanize is based both on mechanistic growth rules as well as a set of weighted probabilities that encourage or inhibit growth. The model is described in detail in Clarke et al. (1996).

Five factors control the behavior of the system. These are: a diffusion factor, which determines the overall outward dispersiveness of the distribution; a breed coefficient, which specifies how likely a newly generated detached settlement is to begin its own growth cycle; a spread coefficient, which controls how much organic expansion occurs from existing settlements, a slope resistance factor, which influences the likelihood of settlement extending up steeper slopes; and a road gravity factor which attracts new settlements toward and along roads. These values, which affect the acceptance level of randomly drawn numbers, are set by the user for every model run.

Four types of growth are defined in the model: spontaneous, diffusive, organic, and road influenced growth. Spontaneous growth occurs when a randomly chosen cell falls nearby an already urbanized cell, simulating the influence urban areas have on their surroundings. Diffusive growth permits the urbanization of cells which are flat enough to be desirable locations for development, even if they do not lie near an already established urban area. Organic growth spreads outward from existing urban centers, representing the tendency of cities to expand. Road influenced growth encourages urbanized cells to develop along the transportation network replicating increased accessibility.

A second level of growth rules, termed self-modification, is prompted by an unusually high or low growth rate. The growth rate is the sum of the four different types of growth defined by the model for each model iteration or "year." The limits of "critical high" and "critical low" kick off an increase or decrease in three of the growth control parameters: diffusion, breed, and spread. The increase to the parameters is by a multiplier greater than one, "boom," imitating the tendency of an expanding system to grow ever more rapidly, while the decrease is by a multiplier less than one, "bust", causing growth to taper off as it does in a depressed or saturated system. However, to prevent uncontrolled exponential growth as the system increases in overall size, the multiplier applied to the factors is slightly decreased in every subsequent growth year.

Other effects of self-modification are an increase in the road gravity factor as the road network enlarges, prompting a wider band of urbanization around the roads, and a decrease in the slope resistance factor as the percentage of land available for development decreases, allowing expansion onto steeper slopes. Under self-modification the parameter values increase most rapidly in the beginning of the growth cycle when there are still many cells available to become urbanized, and decrease as urban density increases in the region and expansion levels off. Without self modification the model produces linear or exponential growth; self-modification was essential for modeling the typical S-curve growth rate of urban expansion.

Since the growth rules in this model are defined primarily by physical factors, the San Francisco Bay area was an ideal test site for this model because of its natural variety: elevations range from sea level to 2500 meters and land use ranges from wilderness to metropolitan areas. Additionally, the region contains diverse patterns of urbanization, reflecting its beginnings in small enclaves clustered around the inland waterway network, the emergence of San Francisco as a transportation hub, and the more recent urbanization of the surrounding valleys, due in part to the improved highway system.

Four major types of data were compiled for this project: land cover, slope, transportation, and protected lands. Six raster image maps of urban extent for the years 1900, 1940, 1954, 1962, 1974, and 1990 were used. 1900 was the seed year, while the other years provided control data against which the model output was compared.

Calibration of the Model

The value of a model's predictions is only as good as the model's ability to be effectively calibrated. Since a model predicting future patterns can only be validated after the fact (when the authors will be unavailable) the validation step is not feasible. The calibration then is possible in only one way, by using the spatial and statistical properties of the past to "predict" the known present. In the HILT modeling effort, we have been using historical land use and satellite data from a variety of sources as calibration points for the model's behavior. At the very least, the model should be able to match the statistical patterns of past growth (within the constraints and limitations of the data) and to provide a probabilistic estimate of today's urban extent that matches reality. Even in the rich data environment in which we have worked, the time limitations of good calibration data are a severe challenge to predictive modeling. Often, a model outcome that performed perfectly statistically made little sense spatially. As a result, we have designed and undertaken a lengthy and computationally-intensive approach to model calibration that we believe to be rigorous.

The basic calibration approach was comparison of the model's output to an historical data set with respect to the key variable, urban areal extent. The Pearson product-moment correlation coefficient (r2) was used as a measure of fit between quantitative and spatial measures. The model's form of implementation evolved through two distinct calibration phases: a visual version which was most useful for broad parameter definition and debugging, and a faster, batch version of the model which was more efficient for generating the large quantity of runs necessary for calibration. The program computes and saves descriptive statistics in designated years for comparison with the control years. The type of output has varied according to the needs of each step of the calibration.

The visual phase first established meaningful ranges of values for the growth parameters through an animation illustrating the impact of change in parameter values. Parameter values were increased by unit increments; each of the final outcomes became a single frame in the animation. Most parameters vary from 0 to 100, necessitating 101 separate runs per variable for the control variables. In each case, all other control variables were held constant at intermediate levels. This ascertained that each control parameter had a unique and controllable impact on the outcome, and it was noticed that some of the variables had clear saturation levels, beyond which increments had relatively little effect. Another benefit of this step was the informal verification that the form and rate of the model's growth was within reasonable bounds.

Two versions of the program with a full set of graphical user-interface tools evolved during this phase. The first was a prototype using the Silicon Graphics tools, which allowed easy animation and display of the resulting images (Figure 1). This version was used in generating animations from the data (Bell et al., 1994; Gaydos et al., 1995). The second version was an X Window System version. The XView toolset was used to create slider bars that allowed adjustment of the critical control parameters while the model was running and the ability to start and stop execution as necessary. This interactive environment provided a way of testing the control parameters' interaction with each other, and for debugging the self modification rules. In addition, a set of measures was explored to allow visual comparison between the actual and the predicted distributions. Symbols positioned on the animated maps showing the actual and predicted centers of gravity for the urban cells proved useful. Also used was the time-sequenced display of a circle with the same area as the predicted distribution drawn at the mean centers of the changing distribution. This allowed rapid visualization of structural changes in the spatial pattern. In addition, urban pixels were color-coded by which cellular automaton rule had invoked their transition.

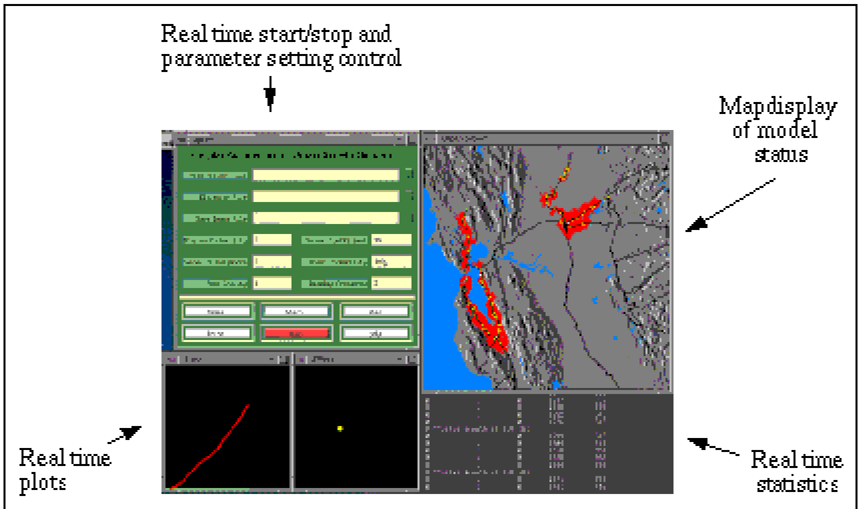


Figure 1: Early interactive version of the model using the Silicon Graphics graphics tools.

Three statistical measures on the form of model growth were displayed on the screen during this phase but were not thoroughly analyzed until the second, graphics-free phase. These measures were the total area converted to urban use, the number of pixels defined as edges, that is with non-urban cell neighbors, which was felt a good measure of the rural-urban fringe nature of dispersed distributions, and the number of separate spreading centers or clusters. The number of spreading centers was computing using image processing algorithms, by eroding the edges of the pattern without disconnecting clusters until each cluster converged on a single pixel, and then counting the remaining pixels, i.e. cluster centers.

During the second phase of calibration, a CPU-optimized batch version of the model was produced, that was faster and more efficient at generating the vast quantity of runs necessary for calibration. Instead of displaying the three statistical measures on the screen (total area, number of edge pixels and number of clusters) the statistics were written into a set of log files for further analysis. An additional statistic became necessary during the non-visual phase, the Lee-Sallee shape index, a measurement of spatial fit between the model's growth and the known urban extent for the control years. This simple measure of shape was adjusted to describe distributions, and was computed by overlaying the observed and the predicted maps of urban extent, computing the union and the intersection of their total areas on a pixel by pixel basis, and then dividing the intersection by the union. For a perfect match, the Lee Sallee measure gives a value of 1.0, and for all others a smaller number, similar to an r-squared value. Inclusion of this statistic proved critical as there was no longer a means of identifying the best spatial match to historical growth now that the visual element was eliminated. These four values were aggregated into a combined score by adding the first three and multiplying by the shape index. The Lee-Sallee value is used as a multiplier because of its importance as a spatial measurement. A separate C program calculates the size, shape, number of clusters, and the Lee-Sallee statistic, as well as correlations between the predicted and observed data.

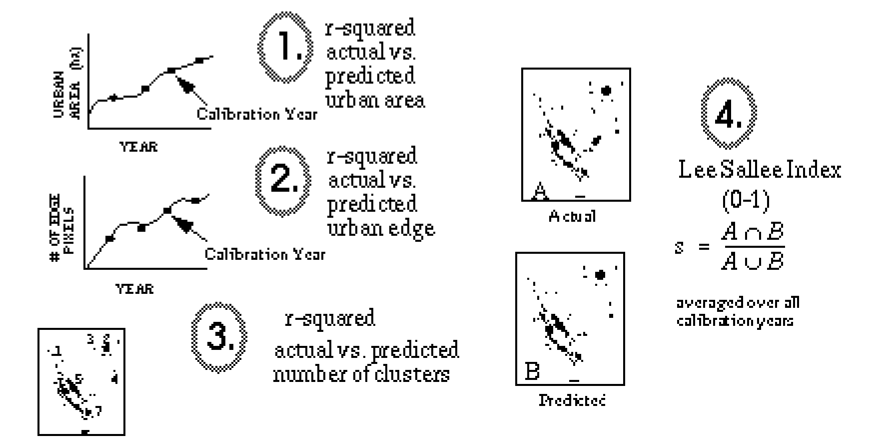


Figure 2: The four calibration measures.

The model requires initialization of five growth parameters and four growth constants before operation, and these were the nine variables adjusted during calibration. The calibration involved finding the best combinations of the five growth parameters which regulate the rate and nature of the types of growth: diffusion, spread, breed, slope resistance, and road gravity. An additional task was definition of the four growth constants which affect self-modification: critical high, critical low, boom, and bust. While these four growth constants will always have the same value at the end of a model run, the growth parameters may have different values from the initial settings if the growth rate has exceeded the critical high or critical low. The calibration strategy was to make minor changes in the nine growth variables, and to compare the variations that resulted in the combined r2 values. A semi-automated phase of the calibration began using shell scripts to first hone in on the best range for each parameter, checking all possible combinations of parameter settings in increments of five, then for a narrower range of parameter values using unit increments.

From the unit increment runs, a sample of 160 different combinations of parameter settings was chosen, or about 10% of the possible number of combinations in the entire set of all possible runs. The statistics from these 160 combinations were examined in detail. To allow for the widely varying outcomes introduced by randomness in the growth process (each run used a different random number generator seed), each individual parameter combination was repeated 10 times, each run with 100 Monte Carlo iterations, creating a sample size of 1000 log files for each of the 160 combinations. The combined r2 value, described above, was used as the basis of comparison between these runs, the purpose being optimization of collective behavior. The statistical package SAS identified the parameter groups with the highest mean and the lowest standard deviation within the repetitions for the combined r2. Histograms of parameter values from the ten best combinations that emerged from this analysis were used to establish the ideal range of values for each parameter. A sensitivity analysis on the range established for each parameter was the final check using the mean and standard error of estimate for each initial parameter setting.

The other form of output from the batch mode of calibration were Monte Carlo probability images produced by the runs. Probabilities were assigned by accumulating over the entire time sequence whether or not each pixel was urbanized. The values then become percentage estimates of probabilities of urbanization in the models target year. Obviously this could be created for any year, and an animated sequence was compiled of the probability maps. These aggregate images allowed comparison of averages over many runs against observed values from the calibration statistics, and allowed the spatial distribution of the outcomes to be studied in probabilistic terms. We experimented with several cartographic variables for the alternative displays of these probabilistic maps.

Evaluation of Calibration

Three sources of error can cause weak correlation during calibration: input error, model error, and parameter error (Debaeke, et al.). A change in the data source used to determine historical extent affected the r2 values edge and cluster. Creation of the historical urban extent layers were assembled from digitized paper maps for years before 1974, while in 1974 and 1990 remotely sensed images were used. The shapes on maps tend to have been generalized by the cartographer while satellite images have far more salt-and-pepper edges. Model error was weeded out early during establishment of parameters during the visual phase. Identical simulations were generated by forcing selection of a known random number seed. Parameter error occurs when the system variables are not independent, and of course this was the case as each parameter influenced more than one growth rule, as well as influencing each other during self-modification. Although a necessary aspect, self-modification introduced an additional level of complexity to the calibration process.

A major advantage of cellular modeling is that all input and output arrays are of a uniform size which permits easy quantitative comparisons between arrays through pixel counts and overlay and direct use of a GIS. Analysis of the output is facilitated as well because cells are of a uniform size and can be readily compared. Furthermore, multiple applications of the model from a variety of starting conditions allows the computation of Monte Carlo-style average aggregate output probabilities of any given cell being urbanized. We can then use the GIS to display, combine, and further analyze the model's predictive maps with other data as layers.

Predictions

Once the model had successfully replicated past urban expansion, it has been used to project future scenarios of urban growth. Because these predictions cannot be verified any time soon, any recommendations based on these results must be understood in terms of probabilities. The predictions are thought to be most valuable as a set of Monte Carlo images, one for each year of prediction that attaches a probability of urbanization to each cell. The dependence of the model on start conditions can be dealt with by starting a future run with the urban map of the present, and by using the ending control parameters for the average optimal historical run. An attractive element of the model is that exclusion layers can be removed, control weakened, and new roads contructed as alternative futures to assist in planning.

Since the model output is a binary array that classifies each cell as urban or non-urban, it can be used to identify areas most at risk of urbanization, or as input to other models that need a layer defining urbanized areas as input, such as a climatology or a land use model. A second level of probability is implicit in these predictions: suffer time decay. As the model moves further into the future, the overall probabilities of predictions decline.

Conclusion

In this work, several important lessons were learned about the process of model calibration in addition to the success of the actual calibration effort itself. In terms of time, the calibration effort has taken as long as the period of model design and construction, and has overlapped productively with the debugging process. For example, only by preparing an animation of the final outcome of parameter range checking could the lack of impact of the parameter due to software bugs be detected, as during a single model run far too many interrelations between variables are taking place simultaneously to detect a lack of response.

The first lesson was that different phases of calibration were best served by radically different versions of the software. During the Model design stage, including initial experiments with small synthetic test data sets and the debugging of the model programs, an interactive graphic version of the model was essential. Literally thousands of graphic displays of model runs were observed while considering the impact of various control factors. We believe that there is no substitute for this "getting to know your data personally" phase. In many cases, the non-spatial nature of hidden or stylized model interface may mask not only errors in the model or data, but prevent the critical experimentation necessary for effective model design.

Secondly, we learned that it is valuable to build the tools necessary to experiment with the model's full range of outcomes. In mathematics or physics, the behavior of a function is first tested by examining behavior at the limits, and no less is true of environmental models. In our case, the range of outcomes came from parameter changes, computing outcomes over the full range of control parameters, by Monte Carlo simulation, and by repetition of each experiment.

Thirdly, we learned that a rigorous mechanism for the evaluation of changing outcomes during calibration is required. The model had four constants and five variables to be assigned values. We used the initial conditions of the past to predict the present because we had key data slices that allowed the computation of measures of the goodness of fit. These measures had to be both statistical averages, and spatial descriptors. Addition of measures of the latter led to the ability to detect spurious outcomes, that although they generated correct numbers, made little sense geographically. We learned the zeroing in on an optimal configuration, in our case a set of initial conditions, required automated test procedures. For this we developed another version of the model and a selection of shell scripts that altered settings and recursively reran the programs, building a data base of outcomes that could be subjected to later statistical analysis. Analysis of these data, stored as log files, required links between the software, the GIS and the statistical package SAS.

Fourthly, we learned that animation is an important cartographic tool for all stages of the model's use, in showing variations in outcomes during debugging, in showing model runs during calibration, and as a tool for communicating the results of the model's predictions to modelers and non-modelers alike after completion of the model. This last lesson may be the most portable of all, that is that modelers are advised to make many spatial representations of their predictions, to use GIS-based cartographic and scientific visualization tools, and to use animation as a way of capturing non-modelers attention with your results.

The model we developed was somewhat special purpose. It was a scale independent cellular model, with some atypical properties such as the use of self-modification. Nevertheless, many of the lessons learned are generic to cellular models, and perhaps to all environmental models. Those working in environmental modeling should recognize the critical nature of calibration as a scientific process, as an opportunity to communicate the model's functions to a broader audience, as the solid ground upon which predictive work stands, and as an opportunity to seek out the limits, statistically and spatially, of a particular environmental model. Calibration should be an integral step in model design, should generate many of the tools for model use, and should build a rigorous scientific base for validation and predictive use of the model.

Acknowledgement

This research was sponsored by the United States Geological Survey under a Joint Research Interchange between Hunter College and the NASA-Ames Research Center (JRI NCC2-5091). Support is gratefully acknowledged. William Acevedo of USGS/Ames is the author of the animations produced by the HILT project.

References

Batty, M. and Xie, Y. (1994a) "Modelling inside GIS: Part 2. Selecting and calibrating urban models using Arc/Info," International Journal of Geographical Information Systems, vol. 8, no. 5, pp. 429-450.

Batty M, and Xie Y, 1994b, "From Cells to Cities" Environment and Planning B 21 S31-S48.

Bell, C., Acevedo, W. and J.T. Buchanan. (1995) "Dynamic mapping of urban regions: Growth of the San Francisco Sacramento region," Proceedings, Urban and Regional Information Systems Association, San Antonio, pp. 723-734. (Appendix 11.3)

Clarke, K. C. Brass, J. A. and Riggan, P. (1995) "A cellular automaton model of wildfire propagation and extinction" Photogrammetric Engineering and Remote Sensing, vol. 60, no. 11, pp. 1355-1367.

Clarke, K.C., Gaydos, L., Hoppen, S., (1996) "A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area," Environment and Planning B. (in press).

Couclelis H, 1985, "Cellular worlds: a framework for modeling micro-macro dynamics" Environment and Planning A 17 585-596

Debaeke, Ph., Loague, K., Green, R.E., 1991, "Statistical and graphical methods for evaluating solute transport models: overview and application," Journal of Contaminant Hydrology 7 51-73

.

Gaydos, L., Acevedo, W. and C. Bell. (1995) "Using animated cartography to illustrate global change," Proceedings of the International Cartographic Association Conference, Barcelona, Spain, International Cartographic Association, pp. 1174-1178.

Kirkby, M.J., Naden, P.S., Burt, T.P., Butcher, D.P. (1987) Computer Simulation in Physical Geography. John Wiley & Sons.

Kirtland, D., Gaydos, L. Clarke, K. DeCola, L., Acevedo, W. and Bell, C. (1994) An Analysis of Human-Induced Land Transformations in the San Francisco Bay/Sacramento Area. World Resources Review, vol. 6, no. 2, pp. 206-217.

Oreskes, N., Shrader-Frechette, K., Belitz, K., (1994) Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science, vol. 263, pp. 641-646.

United States Geological Survey (1994) Human Induced Land Transformations Home Page: http://geo.arc.nasa.gov/usgs/HILTStart http://geo.arc.nasa.gov/usgs/HILTStart

White, R. and Engelen, G. (1992) Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land use patterns, Working Paper no. 9264, Research Institute for Knowledge Systems (RIKS), Maastricht, The Netherlands.

Keith C. Clarke, Department of Geology and Geography, Hunter College- City University of New York, and the City University of New York Graduate School and University Center.

Stacy Hoppen, Graduate Student, Department of Geology and Geography, Hunter College- City University of New York.

Leonard J. Gaydos, US Geological Survey, EROS Data Center, NASA Ames Research Center, Moffett Field, CA.

Roger L. Slothower, Paul A. Schwarz, and Kevin M. Johnston

Some Guidelines For Implementing Spatially Explicit, Individual-Based Ecological Models Within Location-Based Raster GIS.

ABSTRACT

There is growing interest in using raster GIS to model dynamic spatial processes. Moreover, individual-based modeling is an increasingly common aproach to modeling spatially explicit ecological processes. We consider several issues that are critical to implementing individual-based models within raster GIS. These issues are: 1) defining the individual in terms of raster GIS grid cells; 2) defining the spatial neighborhood surrounding the individual; and 3) defining the rules that govern the dynamics of the individual. We also discuss how these issues are used in the modeling of several spatially explicit ecological processes: animal movement, plant competition, and landscape change.

INTRODUCTION

Although many ecologists have embraced GIS for managing, analyzing, and displaying spatial data, fewer ecologists are using GIS for modeling dynamic ecological processes. However, there is growing interest among ecologists and many others in using GIS for dynamic modeling witness this conference and the previous two NCGIA-sponsored conferences. While much of the interest in dynamic modeling using GIS has focused on better integration between simulation models and GIS software (e.g., Hunsaker et al. 1993), there is also growing interest in developing dynamic models within GIS (e.g., Ball 1994, Burrough 1993, Maidment 1993b). GIS and raster GIS software in particular posses many desirable features for developing data-intensive simulation applications, and raster GIS packages are rapidly evolving as generic application development environments for developing spatially explicit models. This paper presents and discusses several issues that are critical to the implementation within raster GIS of a class of ecological models known as individual-based models. In addition, this paper considers how these issues relate to the modeling of three broad classes of spatial processes through examples from the ecological literature: namely animal movement, plant competition, and landscape change. The authors acknowledge that some of these issues have been presented elsewhere, but to the best of authors' knowledge, this is the first summary of these issues and their relevance to modeling within raster GIS.

BACKGROUND

Individual-based models

There are several different approaches to the modeling of spatially explicit ecological phenomena. These approaches, such as reaction-diffusion models and patch models, have been reviewed recently emphasizing plant populations and communities by Czaran and Bartha (1992) and animal populations by Dunning et al. (1995). One modeling approach that has growing interest is the so-called individual-based approach. Individual-based models (IBMs) are organism-based models capable of modeling variation among individuals and interaction between individuals. The use of the individual-based modeling approach by ecologists has become more widespread following the publication of an influential paper by Huston et al (1988). This approach, as discussed by Huston et al. (1988), DeAngelis et al. 1992, and more recently by Judson (1994), acknowledges two fundamental biological principles. The first principle is that individual organisms are behaviorally and physiologically distinct because of genetic and environmental influences. The second principle is that interactions between individuals are inherently localized, i.e., organisms are influenced mostly by nearby organisms.

Although the simulation of many individual organisms can be expensive computationally, often the individual calculations themselves are relatively simple. This simplicity results from the fact that individuals usually interact according to a sequence of basic rules (Huston et al. 1988). These rules, when applied iteratively to many individuals over time, are capable of generating phenomenologically realistic and complex behavior (DeAngelis et al 1986, Huston et al 1988). These relatively simple rules can usually be expressed as algebraic statements which minimize the need for more complex mathematical operations that are associated with other modeling approaches involving differential equations. These simple algebraic statements can be translated readily into the command syntax of many raster GIS packages.

Raster GIS

A raster GIS models space by tessellating it into regular, discrete locations and assigning attributes to each location. The most common tessellation employed by raster GIS software is the square. An individual square cell can be viewed as a unique location within the tessellation or grid. While the grid itself may be georeferenced according to some coordinate system, space within a grid is implicit and is relative to the origin of the grid. Spatial information about a phenomenon is stored for each individual cell, and different phenomena can be stored as separate grids. Most operations in raster GIS process each cell in a grid sequentially and may involve one or more grids.

Ecological Processes

Although the specific spatial processes that are to be simulated are unique to each application, two common types of processes are animal movement and plant competition. Animal movement entails the transfer of an individual organism from one location to another location, and by definition, vacating the previous location. Movement entails both the movement itself as well as the perception or knowledge of space. Other ecological processes, such as foraging and dispersal, often incorporate movement. Plant competition is loosely defined to mean that the presence or proximity of one individual negatively affects the status (e.g. health or probability of success) of another individual plant.

ISSUES

In IBMs, space is continuous and location is explicit, while in raster GIS, space is discrete and location is implicit. Therefore the implementation of an IBM within raster GIS requires translating the definition of individuals, neighborhoods, and rules into the implicit locations used in raster GIS.

Defining the individual

To translate an IBM to raster GIS, it is necessary to define the individual organism in terms of a location. Defining individuals in terms of their locations was described by Molofsky (1994) as the key "trick." In their simplest form, an individual-based model simulates the actions of individual organisms and each organism interacts with others across space and through time. A state vector describing the individual organism includes explicit location coordinates. In the translation to raster GIS, the description of the individual becomes the description of the cell in which the individual organism is found. A state vector describing an individual location contains the necessary attributes to describe the contents of each cell, but in practice, each grid may represent only one attribute of the individual. The state vector now describes a cell, the location of which is implicit in the grid.

One of the defining principles of individual-based models is variation among individuals. When the individual is represented in a cell on a grid as a simple binary presence/absence, then variability between these individuals is either non-existent or must be represented within the rules as a stochastic element. True individuality exists only when the individual is represented by one or more unique values which interact with the rules of the model. It may be that the degree of individuality is determined by the range of values available for the state vector.

There are several approaches for incorporating this variation among individuals in raster GIS models. Increased variability between individuals implies more attributes for each individual, requiring a need to track these additional attributes. Various techniques can be used to keep track of these attributes: a grid for each attribute (resulting in multiple grids); a multiple attribute table that has a record for each individual; placing only similar individuals (such as an age or size class) onto the same grids; or even putting each individual onto its own grid.

It is a logical extension to also consider cases where the 'individual' represents more than one organism (Murdoch 1993). When the individual is a natural group, such as a herd (e.g., Turner 1993), stand, or family, then the rules governing the dynamics are defined in terms of the natural grouping. Aggregations such as populations and communities may not be so readily treated as individuals since the rules of the aggregation may be based on characteristics of the distribution of individuals within the group. However, an aggregation of individuals (especially plants) can often be treated as a uniform or homogeneous location (grid cell) within which the spatial interactions between individual organisms are ignored (Fahrig 1988, Hyman et al. 1991).

Defining individuals in terms of spatial locations is analogous to the modeling of particles using an Eulerian framework. Lagrangian models calculate the trajectories of masses of particles through space and time. Eulerian models, on the other hand, calculate fluxes (or movements) at fixed locations (Maidment 1993a, 1993b). In biological applications, the Lagrangian framework has been used to model the movement of individual organisms (e.g., plankton) while the Eulerian framework has been used to model density fluxes of organisms at fixed locations (Grunbaum 1994).

Defining the neighborhood

In addition to defining the individual, the zone of influence of an individual must also be defined in terms of location. The second principle of IBMs is that interactions between individuals are inherently localized, that is, organisms are influenced mostly by nearby organisms. In the translation of the IBM to raster GIS, the 'localized' area (i.e., the neighborhood) must be defined in terms of the locations which surround a given location. The influences on a location are defined by what is in the space around it.

In IBMs within raster GIS, the simplest neighborhood is one in which only the cells adjacent to the cell being processed are considered. More generally, the neighborhood can be defined as some finite, spatial domain. The definition of neighborhood can include:

- an area defined by its spatial relationship (e.g. adjacency, proximity, direction) with the individual. This area is often a contiguous area defined by a specified shape and distance, but could just as readily be, for example, all locations in a patch 500 meters to the northeast.
- an area defined by characteristics that are independent of location. The neighborhood could be all locations which have the appropriate conditions for growth, regardless of where they are in relation to the individual.
- arbitrarily combined spatial and characteristic requirements, as described in the two previous neighborhood definitions.
- everywhere in the study area. This generally involves consideration of some characteristic which can only be determined by knowing the state of the entire study area. For example, using a cost-surface operation to find the best route for movement requires knowing not only the traits of the path but being able to determine that the path is better than all other possible paths.
- specific locations, without regard for neighboring locations. This is not a spatial interaction, per say, but it is a necessary concept in order for an individual to be able to interact with the characteristics of its own location.

The characteristics of the neighborhood need not be constant and may be a function of time or even of the individual organism. For example, different age classes, represented as separate grids, may require different size neighborhoods for a specific rule.

Defining the rules

The third issue to be considered is the definition of the rules governing an individual's behavior in terms of location. In an IBM, rules describe the dynamics of the individual and are dependent on the definition of both the individual and the neighborhood. These rules may be deterministic or stochastic. In raster GIS, rules are implemented as spatial operations. The types of operations available raster GIS are well described elsewhere (see Berry 1993, Tomlin 1990). The spatial extent of the operations can generally be associated with the classes of neighborhoods described above. A single ecological process may require several spatial rules. Movement, for example, requires at least two rules, one assessing the neighborhood it can perceive (or have knowledge of), and the other actually performing the move.

Time can be modeled as discrete or continuous in a discrete space system. In continuous time, an ecological rule is applied to an arbitrary location and the subsequent application of any rule recognize the results of all previously applied rules at all locations. In discrete time, the rule is applied to all locations 'simultaneously' without recognizing the changes made in adjacent cells. Raster GIS is well suited for discrete time modeling.

One important part of defining rules in raster GIS is an understanding the operations which are currently available. The current technology of raster GIS uses procedures which process each cell of the grid sequentially. Each cell becomes the focus cell, calculations based on an appropriate neighborhood are performed on that focus cell, and the state vector describing the cell is updated. The focus is then moved to the next cell and the calculations are repeated. Consequently, only the state of the focus cell is updated. If the value of a cell is going to change, it can only be changed when that cell is the focus cell. Thus, operations are based on the "focus" of any interaction. This requires a switch from calculating how a cell affects its neighbors to how a "focus" cell is affected by its neighbors. As an example of the two orientations, a tree conceptually disperses seeds to many locations, but raster GIS requires calculating, at each location, the trees that are contributing seeds to that location. This switch requires an inversion of the neighborhood. In an individual organism model, the neighborhood for a seed shadow may be to the southwest of a tree. In the raster GIS, that neighborhood will be inverted to reflect that the major seed input to the focus cell are the trees to the northeast.

It should be emphasized that this "focus" cell restriction may only be a technological limitation of the currently available operations in raster GIS software. The authors believe this to be an artifact of the ease of processing and contributions from the remote sensing field where symmetrical convolution filters were incorporated from electrical engineering (William Philpot, pers.comm.). There may be no reason why processing algorithms could not be written that permit writing to any cell in the grid regardless of the current focus.

DISCUSSION

Movement of animals and competition between plants are two examples of spatial ecological processes which can be modeled as IBMs in raster GIS. A discussion of known applications can provide insight into how the three implementation issues described above are critical to how models are implemented. There are only a few examples of IBMs which model these processes within raster GIS (e.g. Johnston 1992, Rechel 1992). Consequently, this discussion relies on examples from the ecological literature which illustrate the issues presented above.

Some ecological modelers use cellular automata as a modeling approach. Cellular automata are conceptually related to raster GIS in that individuals are placed on a of square grid and the rules are usually implemented within a neighborhood of adjacent cells. Generally, the rules model the transition between states for an individual and are only dependent on the previous state of the individual and the other individuals in the neighborhood, although many of the ecological applications loosen that definition to include a wider range of independent variables (e.g., Ellison and Bedford 1995). Itami (1994) argues that cellular automata models are a logical extension of raster GIS and has developed a specialized GIS for performing cellular automata simulations, as have Theobald and Gross (1994). An alternative to raster GIS and cellular automata is the development of an application-specific system for modeling ecological processes, as seen in the work of Turner et al. (1993). Such systems tend to be application-specific but are often less flexible than the more generic raster GIS packages.

Animal movements

Johnston (1992) builds a predator/prey model which demonstrates that movement includes both a perception neighborhood and a movement neighborhood. The model contains four different sub-models; deer browsing, deer escaping, hunter movement and hunter shooting. The browsing deer has a neighborhood of the adjacent cells in which the rules of interaction (i.e., perception of food resources) and the rules of movement are applied. The hunted deer has a neighborhood in a 3 cell radius (apparently about 37 cells) in which hunters can be perceived. The hunter has a neighborhood of the adjacent cells for perception of the habitat and movement, but has the larger 3 cell radius for perception of deer and shooting.

In a simulation model of foraging behavior, Turner and others (1993) develop an application-specific system in which the individual being modeled is a group of ungulates. With a common perception neighborhood, two different movement neighborhoods are compared; a 1-cell neighborhood and a maximum-forage or maximum-distance-per-day neighborhood. The rules for determining the movement are designed to assess the importance of proximity and quantity of food resources in the four cardinal directions within the perception neighborhood.

Plant competition

In Colasanti and Grime's (1993) cellular automaton model of competition between three generic plant types (ruderal, competitive, and stress tolerator), the individual is defined as a single plant and the neighborhood is composed of the four adjacent cells. The rules model indirect competition as a Markov chain enhanced to include spatial proximity and the resource status of neighboring plants. Silvertown et al. (1992) also use cellular automata to model competitive interactions among five grass species. They also defined simple individuals, simple neighborhoods, and Markovian-type rules for modeling direct competition. In each of these examples, simple models provided significant insight into plant competition.

Landscape change models

Although not defined in terms of individual organisms, landscape change models can be discussed in terms of the issues presented above. Baker (1989) places landscape change models on a continuum of increasing spatial explicitness, culminating in a landscape "element model, in which change in individual landscape elements is modeled" (p. 121). Although Baker explicitly distinguishes these from individual organism models, he acknowledges that they are analogous to individual-based models but from a landscape perspective, that is, the same techniques presented above for individual-based models might also be applied to address questions focused on how landscapes change. The simulation by Turner (1988) of landscape change in Georgia is setup such that cells are arbitrary land uses, the neighborhoods are adjacent cells, and the rules are based on spatially-influenced transition probabilities. Although Turner's individuals are not organisms, the neighborhood and the rules are similar to those in the competition model of Colasanti and Grimes (1992). Ratz (1995) studied the influence of fire on boreal forest succession in which the "individual" was defined as a 4 ha forest stand, the neighborhood was composed of the four adjacent cells, and the rules were governed by stand's susceptibility to burning. While simple in nature, Ratz (1995) concluded that the dynamics of his model were consistent with empirical evidence.

Other considerations

Although raster GIS offers many attractive features as a spatial modeling environment, such as data management, spatial analysis and visualization, it has yet to reach its full potential. Burrough (1992) has argued that the current GIS data structures are too simplistic and that more sophisticated structures are needed. Moreover, the historical emphasis of GIS as a spatial database management system requires that the results of each spatial operation be verified and stored back in the database. Consequently, this imposes severe performance penalties when rapidly updating the database as required for dynamic process modeling. However, the development of GIS software that allows for the updating of grids that are stored in memory and are only stored back to disk when needed may help to reduce this performance penalty. In addition, boundary conditions are an important part of most spatially explicit models because they control edge effects, and the modeling of boundary conditions within raster GIS is often awkward or difficult. Finally, much of the terminology of GIS can be traced to its origins in geography which may not be familiar to ecologists, thus making GIS more difficult to use. Better user interfaces may rectify this thus make GIS packages easier to use. Many of these considerations are being addressed elsewhere during this conference. However, the authors believe that raster GIS is a viable development environment for ecological models because raster GIS is widely used to maintain ecological information, the command syntax is familiar to those ecologists already using GIS, and it includes many spatial operations that are useful for dynamic modeling.

BIBLIOGRAPHY

Baker, William L. (1989) A Review of Models of Landscape Change. Landscape Ecology

2(2):111-133.

Ball, George L. (1994) Ecosystem Modeling with GIS. Environmental Management 18(3):345-349.

Berry, Joseph (1993) Cartographic Modeling: The analytical capabilities of GIS. In M. F. Goodchild, B. O. Parks and L. T. Steyaert (eds.). Environmental Modeling with GIS. New York: Oxford University Press.

Burrough, P. A. (1993) Spatial data quality and error analysis issues: GIS functions and environmental modeling. In Proc. of Second Int. Conf./Wkshp. on Integrating Geographic Information Systems and Environmental Modeling.

Burrough, P. A. (1992) Are GIS data structures too simple minded? Computers and Geosciences 18: 395-400.

Colasanti, R. L. and J. P. Grime. (1983) Resource dynamics and vegetation processes: A deterministic model using two-dimensional cellular automata. Functional Ecology 7: 169-176.

Czaran, T. and S. Bartha (1992) Spatiotemporal dynamic models of plant populations and communities. TREE 7: 38-42.

DeAngelis, D. L. and L. J. Gross, Eds. (1992) Individual-Based Models And Approaches In Ecology. New York: Chapman and Hall.

DeAngelis, D. L., W. M. Post, and C. C. Travis (1986) Positive Feedback in Natural Systems. Berlin: Springer-Verlag.

Dunning, J. B. Jr., D. J. Stewart, B. J. Danielson, B. R. Noon, T. L. Root, R. H. Lamberson, and E. E. Stevens (1995). Spatially explicit population models: current forms and uses. Ecological Applications 5: 3-11.

Ellison, A. M. and B. L. Bedford. (1995) Response of a wetland vascular plant community to disturbance: a simulation study. Ecological Applications 5: 109-123.

Fahrig, L (1988) A general model of populations in patchy habitats. Applied Mathematics and Computation 27: 53-66.

Gao, Peng, Cixiang Zhan and Sudhakar Menon (1993) An overview of Cell Based Modeling with GIS. In M. F. Goodchild, B. O. Parks, and L. DT. Steyaert (eds.) Environmental Modeling with GIS. New York: Oxford University Press.

Grunbaum, D. (1994) Translating stochastic density-dependent individual behavior with sensory constraints to an Eulerian model of animal swarming. J. Math. Biol 33: 139-161.

Hunsaker, C. T., R. A. Nisbet, D. C. Lam, J. A. Browder, W. L. Baker, M. G. Turner, and D. B. Botkin. (1993) Spatial models of ecological systems and processes: the role of GIS. In M. F. Goodchild, B. O. Parks, and L. DT. Steyaert (eds.) Environmental Modeling with GIS. New York: Oxford University Press.

Huston, M., D. L. DeAngelis, and W. M. Post (1988) New computer models unify ecological theory. BioScience 38: 682-691.

Hyman, J. B., J. B. McAninch, and D. L. DeAngelis (1991) An individual-based simulation model of herbivory in a heterogeneous landscape. In M. G. Turner and R. H. Gardner (eds.), Quantitative Methods in Landscape Ecology. New York: Springer-Verlag.

Itami, Robert M. (1994) Simulating spatial dynamics: cellular automata theory. Landscape and Urban Planning 30:27-47.

Johnston, Kevin M. (1992) Using Statistical Regression Analysis To Build Three Prototype GIS Wildlife Models. GIS/LIS '92. 374-386.

Judson, O. P. (1994) The rise of the individual-based model in ecology. TREE 9: 9-14.

Maidment, David R. (1993a) GIS and hydrological modeling. In M. F. Goodchild, B. O. Parks, and L. DT. Steyaert (eds.) Environmental Modeling with GIS. New York: Oxford University Press.

Maidment, David R. (1993b) Environmental Modeling Within GIS. In Proc. of Second Int. Conf./Wkshp. on Integrating Geographic Information Systems and Environmental Modeling.

Molofsky, J. (1995) Population dynamics and pattern formation in theoretical populations. Ecology 75: 30-39.

Murdoch, W. W. (1993) Individual-based models for predicting the effects of global change. In P. M. Kareiva, J. G. Kingsolver, and R. B. Huey (eds.) Biotic Interactions and Global Change. Sunderland, MA: Sinauer Assoc, Inc.

Ratz, A. (1995) Long-term spatial patterns created by fire: a model oriented towards boreal forests. Int. J. Wildland Fire 5: 25-34.

Rechel, J. L. (1992) Geographic information systems modelling of wildlife movements across fire and urban disturbed landscapes. Bull. Ecol. Soc. Am. 73 (2 Suppl.): 316.

Silvertown, J., S. Holtier, J. Johnson, and P. Dale. (1992) Cellular automaton models of interspecific competition for space -- the effect of pattern on process. J. of Ecology 80: 527-534.

Theobald, D. M. and M. D. Gross. (1994) EML: a modeling environment for exploring landscape dynamics. Comput, Environ. and Urban Systems 18: 193-204.

Tomlin, C. Dana. (1990) Geographic Information Systems and Cartographic Modeling. Englewood Cliffs, NJ: Prentice Hall.

Turner, Monica G., Yegang Wu, William H. Romme, and Linda L. Wallace (1993) A Landscape Simulation Model Of Winter Foraging By Large Ungulates. Ecological Modelling 69:163-184.

Turner, M. G. (1988) A spatial simulation model of landuse changes in a Piedmont County in Georgia. Applied Mathematics and Computation 27: 39-51.

AUTHORS

Roger L Slothower
Department of Natural Resources
Cornell University
Ithaca, NY 14853
rs26@cornell.edu

Paul A. Schwarz
Cornell Theory Center
Ithaca, NY 14853-3801
schwarz@tc.cornell.edu

Kevin M. Johnston
Environmental Systems Research Institute, Inc
Oakham, MA 01068
kjohnston@esri.com

SF23 1/5/96 1600

Mark Gahegan

Visualisation Strategies for Exploratory Spatial Analysis

Abstract

The use of computer visualisation as a means to analyse complex geographic datasets is discussed. Visualisation is a valuable tool for conducting exploratory data analysis on geographical data, making good use of the human eye's unparalleled ability to recognise inherent structure and relationships. Existing Geographic Information Systems are extremely poor at visualisation, being limited to 2D (2.5D), with a very restricted set of visual attributes with which to convey information (position and colour). The use of more sophisticated tools is discussed in detail. Specifically, two different approaches to the visualisation of complex environmental datasets are described. The first approach is designed around the concept of data spaces, and presents the user with a series of 'pictures', each describing a data space, and with tools to explore the synergy between data spaces. The second approach uses knowledge-based visual composition, where the most salient properties in the data are assigned to the most striking visual attributes. This approach uses visual data combination techniques to merge data layers into a single 'picture' where possible.

Results from the application of both of these approaches on datasets involving several multi-dimensional thematic layers of environmental data are presented. The relative merits of each as a tool for exploratory spatial analysis on complex geographical datasets are discussed.

Introduction

The aim of this paper is to provide a comparison of two substantially different approaches to visualisation; both specifically aimed at large and complex geographic datasets. To allow a more objective comparison, both approaches are applied to the same publicly available NASA Pathfinder dataset, from the Kioloa region in New South Wales, Australia (Kioloa, 1994).

Visual Data Analysis

With vast amounts of data now being available in digital form, there has been a shift away from deductive approaches, using small amounts of data, towards an inductive approach where models are not so much hypothesised as observed from the data itself (Tang, 1992). When attempting to analyse or model physical processes it is usually the case that we cannot fully describe them in a deterministic manner. This is due to the inherent complexity in these processes, and also to inappropriate, inaccurate and missing data. Exploratory data analysis is a useful tool for providing insight into these complex and subtle relationships within the data.

Visualisation offers a means to combine and view several channels (attributes) of data concurrently. It is being used here as a technique for Exploratory Data Analysis; indeed, the term Visual Data Analysis has been coined to describe the use of visual techniques for this purpose (Warner, 1990). The approaches to visual exploratory analysis described here are largely hypothesis free.

It is not the aim of visual analysis to *recognise* structure in data, but rather to present the data to the user in such a way that the task of interpretation is made easier. As such its role differs from that of statistical approaches, which provide summary information and from processing approaches, which transform or classify the data in some manner. The visualisation approaches presented aim to minimise processing on the data. Transformation performed on data often change some of the inherent relationships. Since these relationships may be of importance, we try to preserve them wherever possible. Transformed or summarised data may of course be visualised if it is available, but it is not the aim of visualisation to produce it.

For many tasks involved with image interpretation, the human visual system has unparalleled performance, so rather than attempting to automate the process, we instead try to facilitate it. In order to achieve this, data must be presented so that visual emphasis is given to the phenomenon under investigation; that is, the *target* is characterised in a visual *stimulus space* by unique properties that may be distinguished from other distracting properties (Csinger, 1992). State of the art visualisation environments provide a fine level of control over the stimulus space, which can be used to promote visual emphasis (pop-out) and to enable more data to be included within a visualisation, making use of properties such as colour, texture and orientation. Possible aims of visual analysis are:

1. To search for trends in the data.
2. To examine the variation of one data channel over another (correlation).
3. To discover combinations of variables, across several datasets that together offer a means of differentiating some target or concept.

One might assume that a Geographic Information System (GIS), specifically designed for the storage, manipulation and display of geographic data, would provide a useful environment within which to explore the data visually. But unfortunately in current GIS, the number of graphical primitives under the direct control of the user is small. On most systems, only 2D or 2.5D is supported, along with a choice of colour. Given the case of position (x, y) and colour, we have only three visual attributes that can be used for display. In contrast, the number of attributes within the data may be very large. Rather than extending an existing GIS (Hartmann, 1992), or building specific tools (Haslett *et al*. 1991), we have chosen to exploit existing visualisation environments, specifically IRIS Explorer. IRIS Explorer was developed by Silicon Graphics as a visualisation system using the Inventor Geometry Engine. It is now developed and marketed by NAG (UK).

The Visualisation Problem

The problem of visualisation can be broken down into two independent issues. The first issue concerns the overall complexity and volume of the data; specifically, there are often more data available than can be simultaneously visualised by even the most advanced environments, so it becomes necessary to select some channels whilst rejecting others. The second issue concerns the assignment of data to the parameters that control the visualisation, as determined by the environment chosen. Without a structured approach, the search space of possible visualisations can soon become prohibitively large.

The automated production of visualisations requires the application of knowledge concerning data reduction and the assignment of data to visual parameters. Additionally, a number of different types of constraints must be applied concerning properties of the data and the facilities provided by the visualisation environment. Various researchers have concentrated on the need to introduce a structure to manner in which visualisations are constructed (Beshers and Feiner 1993, Robertson 1990, Rogowitz and Treinish 1993). The composition approach builds on the work of Senay and Ignatius (1991, 1994) making use of three distinct types of knowledge (i) knowledge of the data, (ii) knowledge of computer graphics and visualisation tools and (iii), knowledge of visual perception. This provides a much needed structure to a highly underconstrained problem.

Alternative Architectures for Visualisation

The two alternative approaches to visualisation taken here are: (i) Visualisation of Data Spaces and (ii) Visual Composition, controlled by various types of expert knowledge.

The architectures of the two approaches are shown schematically in Figure 1 and Figure 4. The visualisation environment is the software that provides the graphical facilities required. Example environments are IRIS Explorer, ARC/INFO and Adobe Postscript. Each supported environment requires its own map builder, to generate valid visualisation schemes in the language or script of the environment (for example an Explorer Map). Additionally, each environment imposes its own restrictions on the ways in which data may be visually combined. The environment is described in a common format that is used to ensure the production of valid visualisation schemas; that is, using only the facilities provided by the environment. The environment dependent components are shown isolated by a grey dashed line, representing the extents to which they influence the system as a whole. By encapsulating the environmental constraints in this way, the task of changing to another environment is simplified. Constraints imposed by the data, the task and the perceptual knowledge base are isolated from those that are system dependent. To illustrate the point, it is often desirable to take the same datasets into the field on a portable Notebook PC, running *ArcView* as opposed to *Iris Explorer*. Only the system dependent modules need be re-written. Obviously, the resulting visualisations will be more restricted, due to the reduction in dimensionality of the stimulus space. However, the same rules and knowledge still hold.

The key element in both schemas is the assignment module. This takes in channels of data and assigns them to the various visual attributes that the environment supports. For the data space approach, the assignment is made according to user judgement or underlying correlation within the data; for the composition approach it is made according to captured knowledge.

As discussed in the introduction, it is not the aim of visualisation to *process* the data. The one necessary exception concerns the *scaling* of data. Since both approaches display several channels of data simultaneously, we must ensure that the resulting axes are similar in length. There is an implicit assumption that data may be scaled as required, without affecting its dynamic range.

The Data Space Approach

Data spaces are formed by assembling data channels into graph structures, which often have a traditional theme. Examples, as given by Lees (1994) are: spatial (real), spectral, environmental and taxonomic. The components of these data spaces are not specified, but rely on the judgement of the user. The environmental data space could encompass rainfall, humidity, temperature, soil pH, etc.. As pointed out by Aspinall and Lees (1994), the relationships between these data spaces are complex: moving a small distance in environmental space may correspond to a large movement in real space, and vice versa. The user may interact with a data space by changing the viewing angle in three dimensions and by highlighting a point or region in one space- the system can then highlight the corresponding region(s) in another space. Figure 1 shows the system architecture.

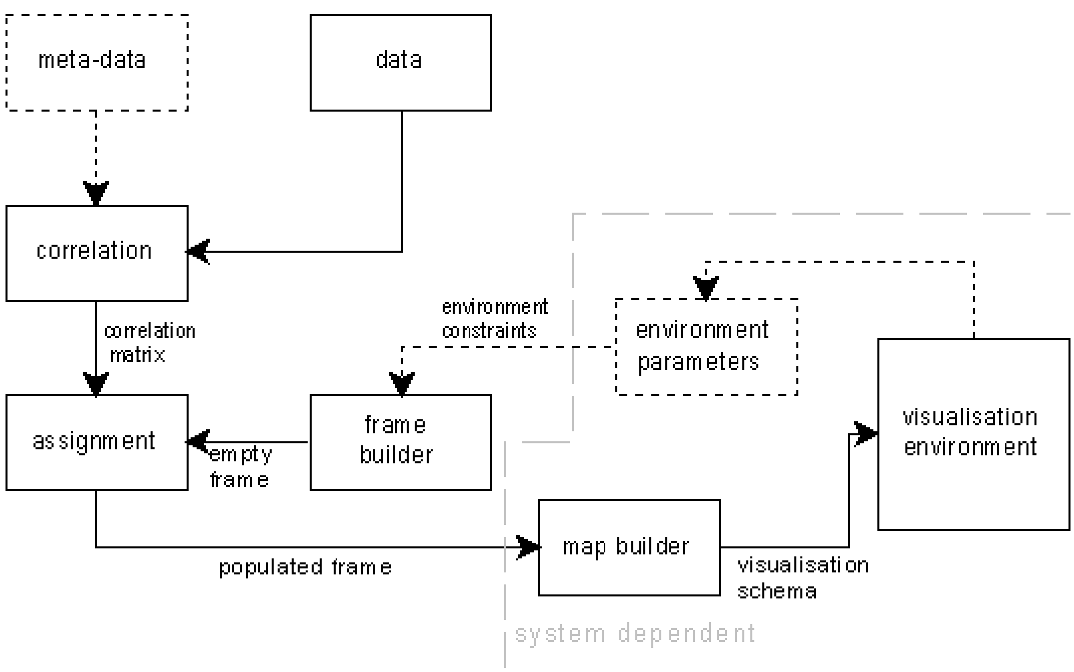


Figure 1: Architecture for Data Space Visualisation

Examples

In the spaces constructed for visualisation, only the familiar visual attributes of position (x, y, z) and colour (c) are used. A maximum of three axes are constructed, this being the maximum number of dimensions that can be displayed without projection in graph form. In the example results given (Figures 2-3), each rendered sphere represents a training site for which all dataset values can be gathered and whose (predominant) landcover class is known. Figure 2 represents a data space with a hydrological theme, showing flow length, flow accumulation and slope.

The user is able to study the problems of class separability within and between the data spaces. To aid in this activity, the colours may be fixed throughout all the data spaces, giving the user a means of tracking data from one space to another. The stimulus space over *n* data spaces is defined as: $S = \{c, x1, y1, z1, x2, y2, z2, ..., xn, yn, zn\}$.

This type of approach may be entirely generic, or may contain a *task* of the form "*discriminate A* (*from B)"*. Both *A* and *B* may be one or more classes of data. (For tasks such as classification, it is often desirable to group together data channels that provide the most discrimination of some target concept, in this case the desired output classes.) To aid this activity, the data spaces can be automatically configured, using analysis tools to search for correlation among the data channels. Figure 3 is an example, showing the use of automated techniques to provide the assignment of data channels to data spaces. Automatically generated spaces may not have a traditional theme as such, so care must be taken when interpreting the data since the axes of the spaces may represent substantially different concepts. Figure 3 shows a data space automatically configured using cluster analysis to emphasise the *Rainforest Ecotone* training class from the Kioloa Dataset. The result is a spectral data space of Landsat bands 2, 4 and 7. The inseparability between *Rainforest* and *Rainforest Ecotone*, shown in red and yellow respectively, is clearly visible; as is the differentiation of *Paddock*, shown as blue. This kind of information is helpful when designing a classification scheme.

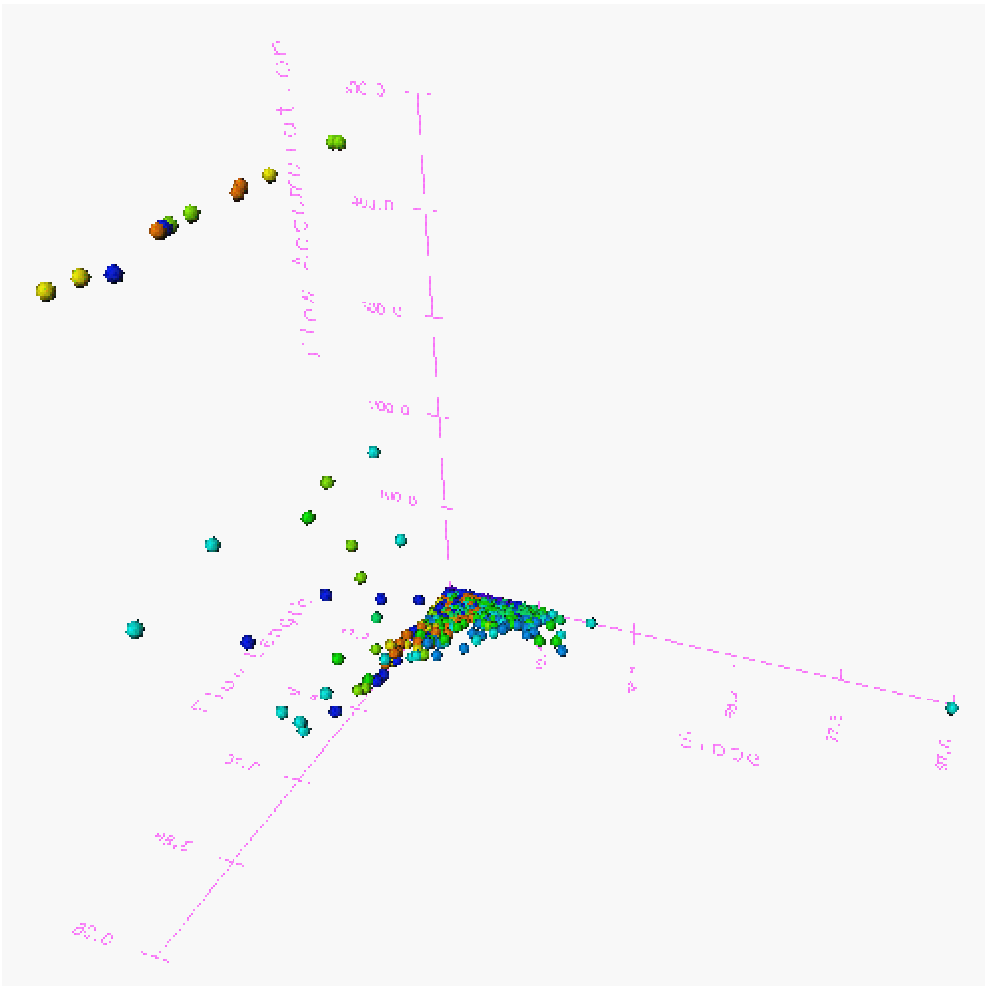


Figure 2: Hydrological data space showing slope, flow length and accumulation

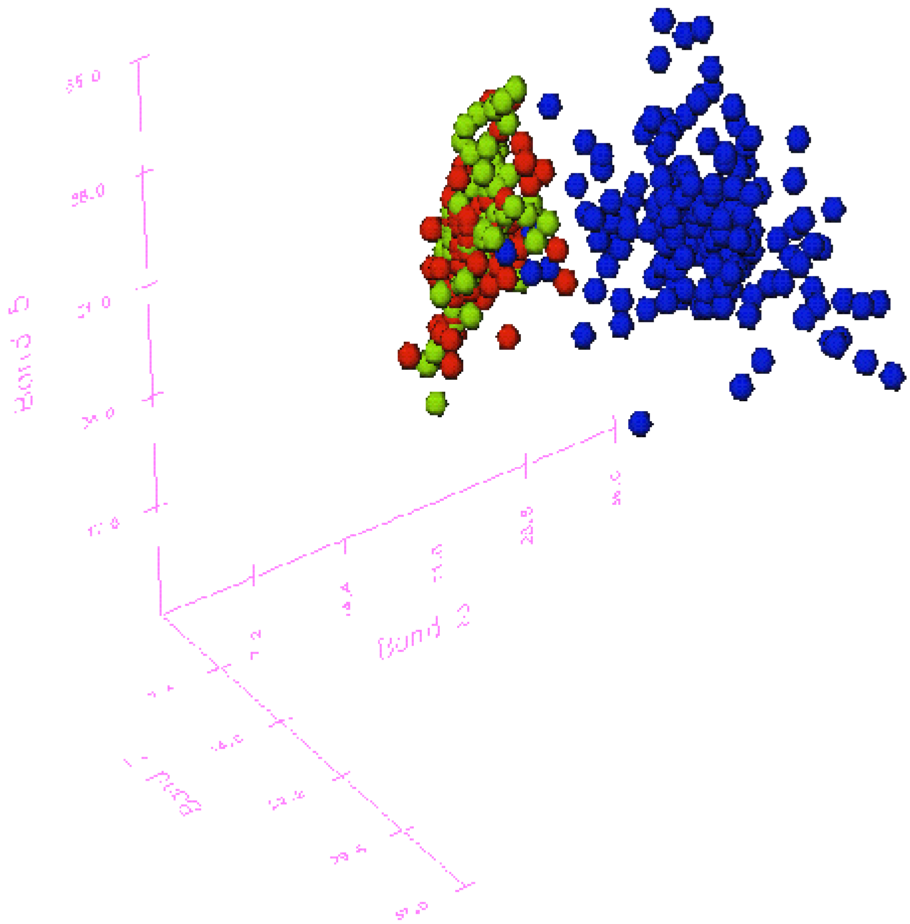


Figure 3: Spectral data space of Landsat bands 2, 5 and 7, formed by cluster analysis and showing *paddocks* (blue), *rainforest* (yellow) and *rainforest ecotone* (red).

Configuration of Data Spaces

The construction of data spaces is a four stage process:

1. Assignment of channels to data spaces. This can be done by an expert, a knowledge base or in an automated fashion using techniques to maximise (minimise) correlation between data channels. Techniques used so far are: simulated annealing, cluster analysis and multi-dimensional scaling. The results obtained by one method are not necessarily consistent with another.
2. Selection of channels to visualise. Since visualisation is restricted to three channels per data space, channel reduction may be required. Automated techniques can again be used, for example principle component techniques will identify the three channels which together account for the most variance in the data. Alternatively, the user may decide.
3. Scaling of data. To ensure that the graph axes are similar in length, channels of data are individually scaled. Again, various techniques may be used, including histogram equalisation.
4. Rendering. The resulting data spaces are rendered for the user to explore.

The Visual Composition Approach

This approach is more complex, requiring first that the user supply a specific *task*. The rationale is that the utility of data depends very much on the task at hand; a channel that is highly relevant to one task may become just a source of noise for another. Tasks are of the form: < *task property domain* >.

Tasks are the basic instruction and include correlate, explore and search (identification). *Properties* are used to define specific *goals* for which the task is to be constructed. These represent a conceptual *target* to which the visualisation is directed. Properties will change according to the application. For the Kioloa dataset, possible tasks include floristic classification and landform analysis. The selection of data channels to be used and their relative importance varies widely with these properties. *Domains* offer a method of loosely grouping data, so that the task is directed at a particular region in the universal data space. Domains are similar to the concept of *views* in relational databases, providing a means of viewing the data that is suitable for some tasks. The domains might be chosen to represent seasonal changes if multi-temporal data is available. The user can issue specific, highly focussed requests, such as: < *explore floristics spring* >.

An overview of the architecture is shown in Figure 4. A rich set of graphical primitives (marks) is used. Examples include surfaces, lattices, arrows and spheres. Each of these has its own set of visual attributes, which include positional variables (x, y, z), colour variables (red, green, blue) and may include others such as direction and opacity. Each layer of marks contains as variables a set of r visual properties that may be used to control the appearance of (up to) r channels of data, so the visual stimulus defined by each layer is: $S = \{v1, v2, ..., vr\}$.

Two or three of these variables will be *bound* to control position, the others will be *free*. Typically, a GIS provides only one layer of 2D marks containing only one free visual property (colour). In contrast, the total stimulus space for a compositional visualisation is given by the sum of the stimulus over all layers used.

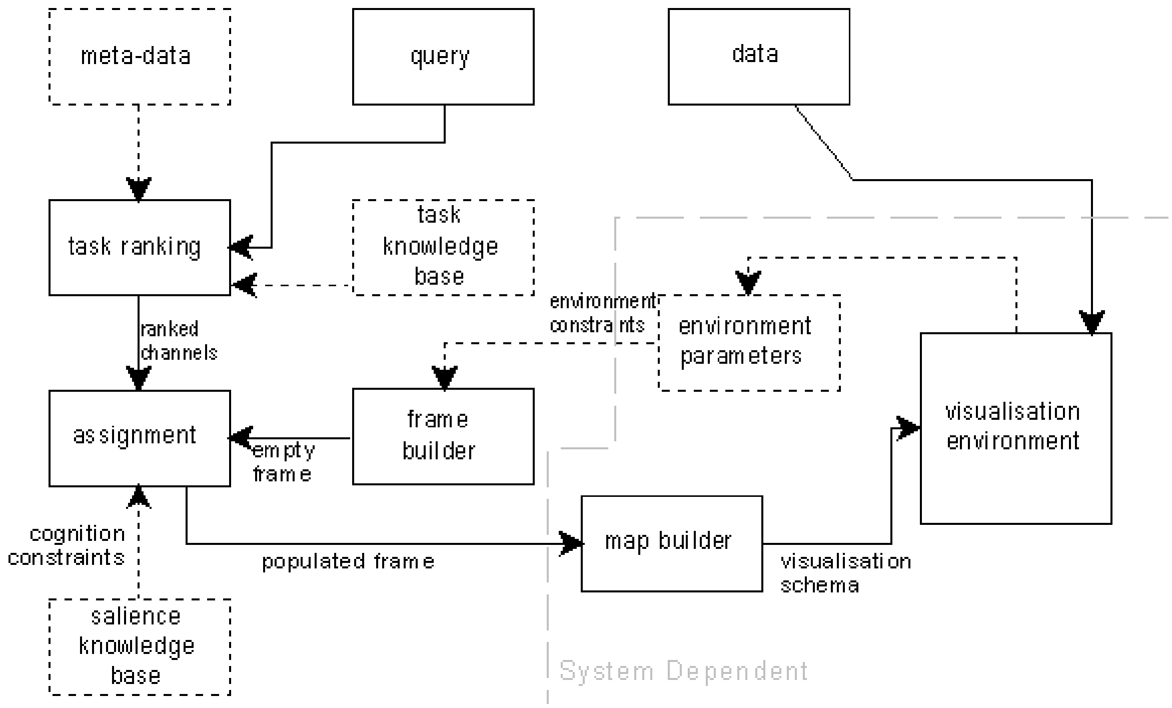


Figure 4: Architecture for Compositional Visualisation

Examples

Figures 5 and 6 show visualisations with a floristic theme. The lower surface is constructed by a combination of four channels of data controlling red, green, blue (from Landsat) and vertical displacement (from DEM), and sharing the same positional variables. An additional layer in the form of arrows is used to encode a further six values at selected sites. Each of the arrows represents a ground truth sample point; the origin indicates position and the colour denotes the predominant class as follows: *Dry Sclerophyll* (Red): *E. Botroyoides* (Orange): *Lower Slope Wet* (Yellow): *Wet E. Maculata* (Kermit Green): *Dry E. Maculata* (Moss Green): *Rainforest Ecotone* (Light Blue): *Rainforest* (Blue): *Paddocks* (Purple): *Water* (Pink). In Figure 5 the arrow direction is controlled by Landsat bands 2, 4 and 7 (x, y and z displacement respectively) at the sample points. The realisation has a high visual impact and reveals some interesting insight into the spectral data, notably: (i) sample points from the same class are not always characterised in a similar way (arrows of the same colour have differing orientation); (ii) sample points from different classes may be characterised in a similar way (arrows of differing colour with similar orientation). We may safely conclude that regions with the same landcover class are not always distinguishable using spectral data alone. Figure 6 uses an assignment of channels to arrows as Landsat 7, geology and surface accumulation to x, y and z displacement respectively. Again, the ability of these channels to discriminate between the training classes is shown by the orientation of the arrows.

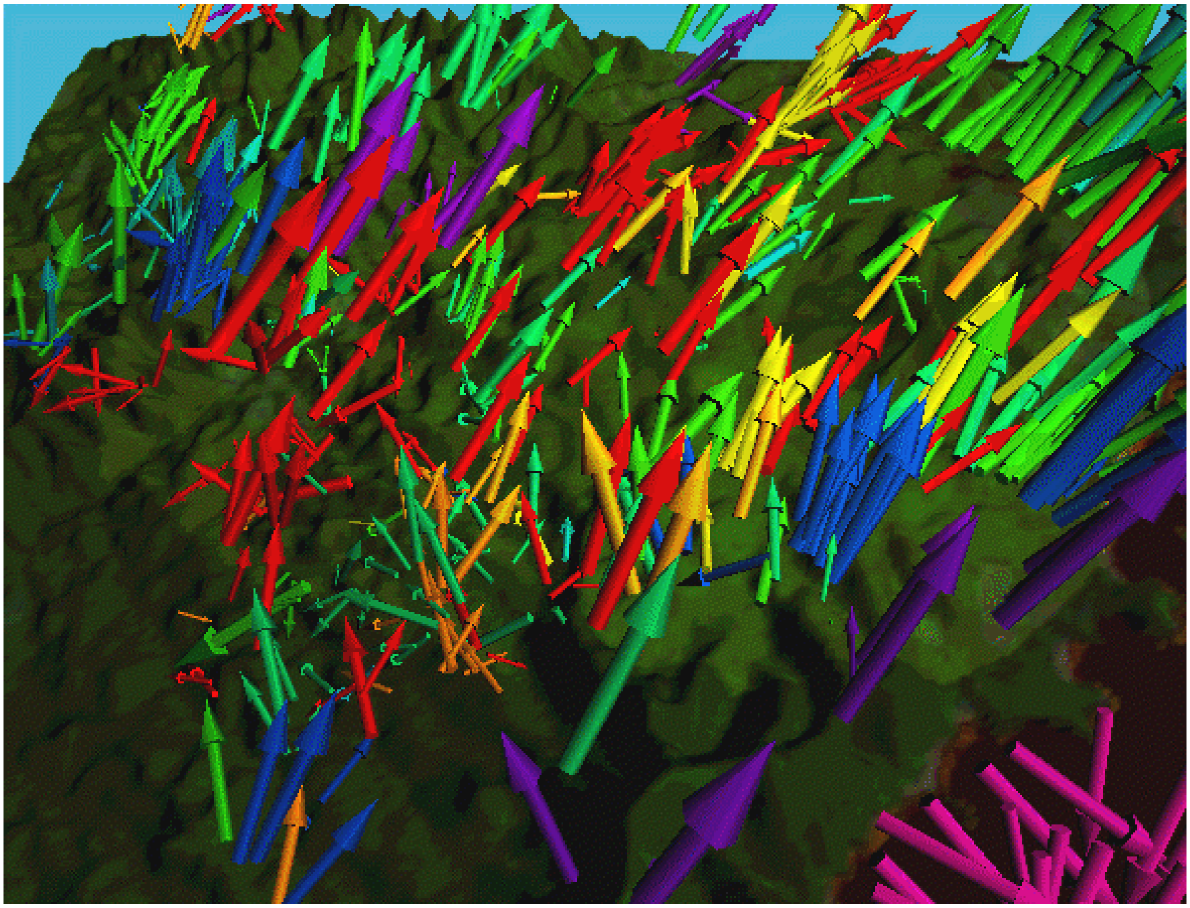


Figure 5. Topographic surface with superimposed arrows showing spectral response at ground truth sites.

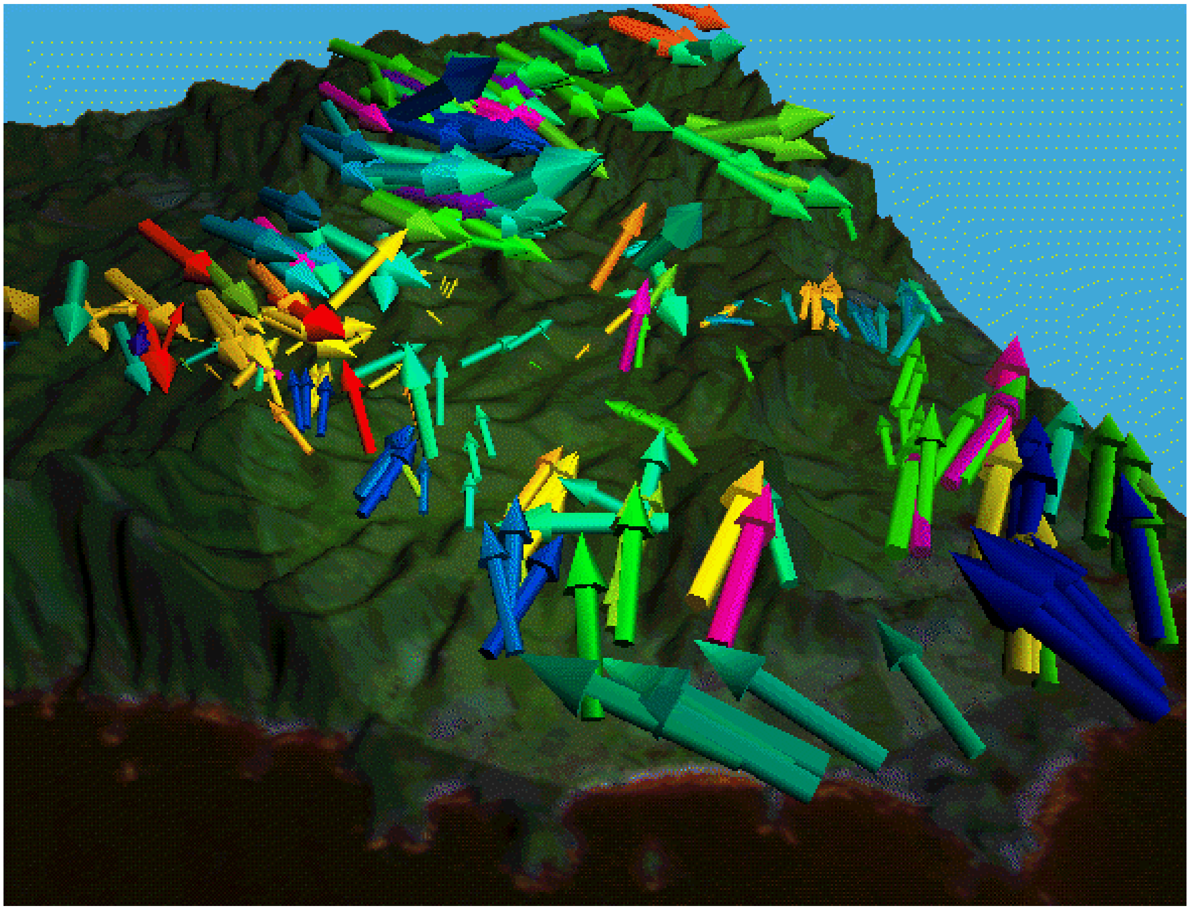


Figure 6. Topographic surface with superimposed arrows showing Landsat 7, geology and accumulation response at ground truth sites.

Figure 7 shows a visualisation constructed using two surfaces which are then intersected by reducing the vertical displacement separating them. The result is that one surface then protrudes into the other. In this case, the upper surface is defined again by Landsat bands (2, 4, 5) and elevation. The lower surface z displacement is provided by Landsat band 7, and the colour from a hydrological accumulation map where colours range from green, signifying low surface accumulation to blue, signifying high accumulation. Structural relationships are evident in that the lower surface protrudes into the upper surface along the valleys. Figure 7 is in fact just one frame from an animation sequence of the two surfaces being intersected (URL is given below).

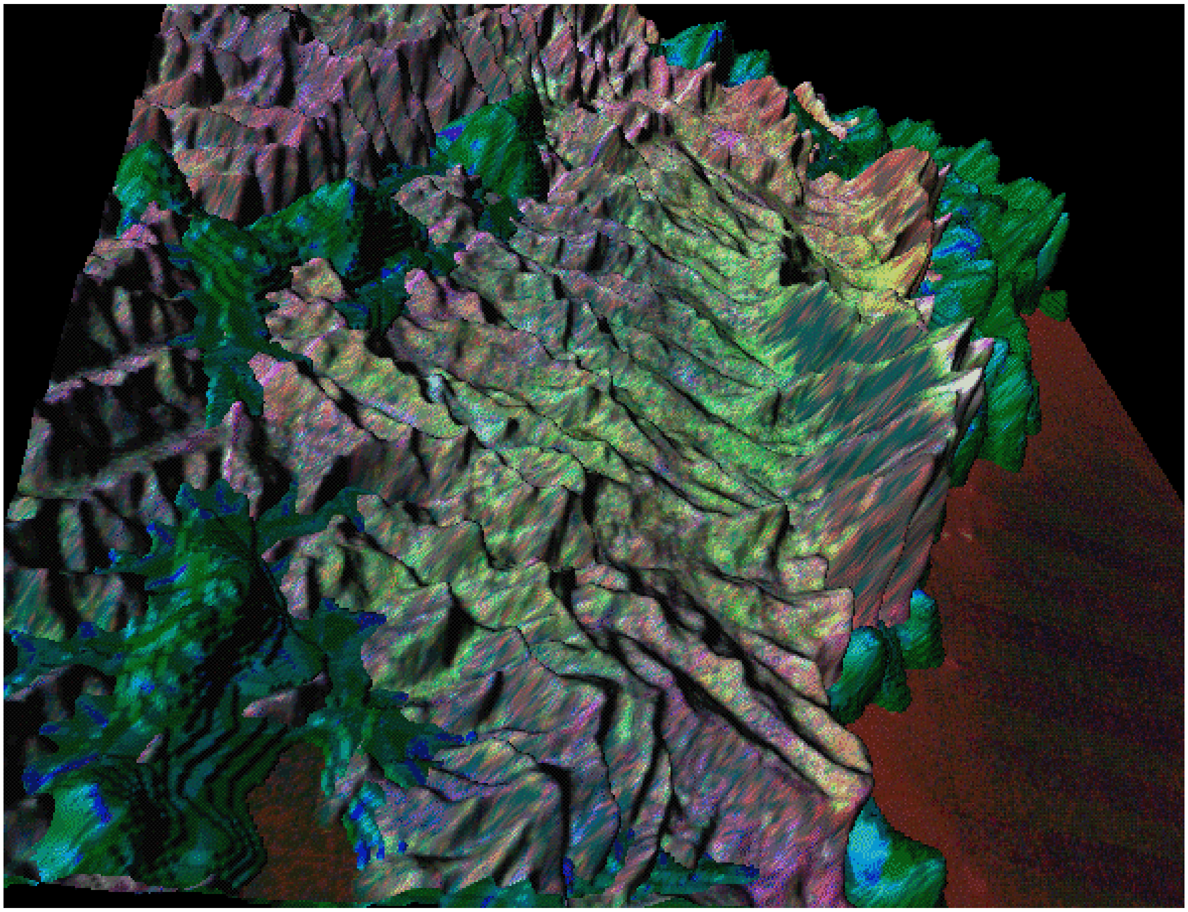


Figure 7. Elevation surface intersected with Landsat band 7 surface.

Higher quality versions of these images are available remotely:
Figure 5 (675 KB) Hi Res interactive VRML Scene (192 KB)
Low Res interactive VRML Scene (30 KB)
Figure 6 (1 MB)
Figure 7 (863 KB)
QuickTime Movie (2.1 MB)

Construction of Compositional Visualisation

The construction of the visualisation is a six stage process with the following steps:

1. Ranking of channels. Data channels are ranked according to their utility to the given task. A task knowledge base, configured by a domain expert, is used to provide the ranking. Channels with little or no utility are discarded.
2. Selection of graphical primitives (marks). The type of mark used depends on various aspects of the data, such as its dimensionality.
3. Assignment of mark visual attributes to channels. This is a state-space search during which the various visual attributes available in the mark descriptions are assigned to the selected channels on a cost minimisation basis. Assignment is controlled again by metadata, such as the statistical type of the data and by knowledge of human cognition

from the salience knowledge base. Two operators are defined, as given by Senay and Ignatious (1991): (i) *Mark Composition*, where spare capacity in marks is absorbed by assigning multiple channels of data to the one set of marks, sharing the same geometry (ii) Axis Composition, where new layers of marks are aligned vertically and share the same axes (usually real space). Mark composition is favoured over axis composition since there is a high cost in terms of cognition when new layers are introduced.

4. Scaling of data. Data requires scaling so that the visual impact of the required concepts is promoted.
5. Rendering. The visualisation is rendered for the user to explore.
6. Feedback from the user. The user may rank the resulting visualisation according to its perceived usefulness for the required task. Feedback is used to change the relative importance of the various heuristics that control the behaviour of the assignment process.

The implementation of this approach is coded using CLIPS, an expert system shell. A full description of the approach is given in Gahegan and O'Brien (1996).

Discussion of Results

In an ideal inductive system, all pre-defined structure that may be imposed on the data would be eliminated, but in practice this is not always possible or desirable. In the data spaces approach, some global assumptions (or deductions) are made regarding the groupings that should be applied to different datasets. In the compositional approach, some global assumptions are made regarding the salience of different datasets to specific tasks.

Methods based on Data Spaces

Data spaces are a simple concept requiring little of a visualisation environment. They are easy to configure and do not rely heavily on supplied knowledge. Since they use only familiar modes of presentation they are conceptually undemanding to use. Expertise is required when defining the data spaces initially.

In terms of visualisation, data spaces may contain a high degree of redundancy since many visual attributes are not used. Because of restrictions on the dimensionality of spaces, some data channels may also be omitted which may mean that certain relationships go undetected. Whilst an analytical model is not explicitly imposed, the idea of pre-defined data spaces leans towards a deductive approach. However, it may be necessary for some level of structure to be imposed on the data in order that the outcome can be comprehended visually by a user. Additionally, the individual data spaces have a high degree of visual separation, requiring a good deal of effort on behalf of the user to study effects between data spaces.

Methods based on Visual Composition

Visual Composition requires a fixed concept (real space in the examples shown) over which layer and mark composition may be applied. It also requires the use of knowledge bases and an advanced visualisation environment. The search space of potential visualisations is large and must be carefully managed. Some poor visualisations are produced as a result of incomplete knowledge capture, and the complexity of modelling human cognition. Structure is imposed on the data by the task model, and this limits to some extent the types of renderings that may be produced. Again this leans towards a more deductive approach, but it is still possible to explore unanticipated structure since little is assumed regarding likely cause and effect within the data prior to visualisation.

The renderings are artificial, and hence harder to comprehend initially. This problem appears to be eased by adopting a landscape paradigm by constructing surfaces in the spatial domain (Robertson, 1990), which displays the data in a more familiar form. The visual separation of data is minimised as far as possible, by mark composition so that a low cognitive effort is required to study inter-relationships. However, there can be a high cognitive load in orthogonalising the various visual attributes and relating changes in visual appearance to data values.

The compositional approach described here was developed specifically to support the needs of mineral exploration geologists, providing visualisations for tasks such as < *explore geology shallow* >. In this field, it is common to use in excess of ten channels of data for some analyses and the approach has proved a valuable tool for exploring structure simultaneously across large exploration datasets. Visualisation at Curtin

Conclusions and Future Directions

The simple concepts used in the data space approach make it ideal for preliminary visual analysis, or for the inexperienced user. By contrast the composition approach is harder to grasp initially, but can provide a more focussed means of viewing the data once a task is established. Structural relationships over space are more easily observed in the composition approach, due to both the strong emphasis on positional information (it is fixed to allow axis composition), and to the coercion of several channels of data into a single layer. There is no reason why a visualisation strategy should not make use of both of these approaches, and it is not intended that the user be forced to make a choice.

Whilst it would be desirable to have sophisticated visualisation tools available from within GIS, it is a fairly simple matter to pass data from the GIS to the visualisation environment, since integration occurs at the external level within both systems. In the long term however, it may prove highly beneficial for full integration to occur. Apart from the issue of a single environment being easier to use, GIS operations could then be animated to allow the user to study in detail their effects on the data.

Future research will concentrate on two areas. Firstly, it is hoped to move even further towards a truly inductive approach, so automated techniques to induce structure from the underlying data will be explored further. Secondly, the visualisations produced will be used as a basis for the construction of classifiers. If good visual separation of concepts can be achieved, then this should be directly translatable into hybrid classification schemes that may then be applied to the data.

Acknowledgments

The work reported is based on visualisation research projects being undertaken at Curtin University, Department of GIS. The author particularly wishes to acknowledge Darren Christopherson and David O'Brien for providing some of the figures used in this paper. Thanks go to Brian Lees (ANU Dept. of Geography) for supplying the data.

References

Aspinall, R. J. and Lees, B. G. (1994) Sampling and Analysis of Spatial Environmental Data. *Proceedings, 6th International Symposium on Spatial Data Handling, Ed: Waugh T C and Healey R G,* Edinburgh, Scotland, pp. 1086-1098.

Beshers, C. and Feiner, S. (1993) AutoVisual: Rule-Based Design of Interactive Multivariate Visualisations. *IEEE Computer Graphics and Applications,* pp. 41-49.

Csinger, A. (1992) The Psychology of Visualisation. *Technical Report Series*, Department of Computer Science, University of British Columbia.

Gahegan, M. and O'Brien, D. (1996) Strategies for the Visualisation of Complex Geographical Datasets. To appear in: *Spatial Computing*, Scientific Press.

Hartmann, J. L. (1992) Visualisation Techniques in GIS. *Proceedings, EGIS '92. Vol. 1*, EGIS Foundation, Utrecht, Netherlands, pp. 406-412

Haslett, J., Bradley, R., Craig, P., Unwin, A. and Wills, G. (1991) Dynamic Graphics for Exploring Spatial Data with Application to Locating Global and Local Anomalies. *The American Statistician*, Vol. 45, No. 3, pp. 234-242.

Kioloa, (1994) The ANU Kioloa Dataset: Distribution Documentation. *Department of Geography, Australian National University*, Canberra, Australia.

Lees, B. G. (1994) Decision Trees, Artificial Neural Networks and Genetic Algorithms for Classification of Remotely-Sensed and Ancillary Data. *Proceedings, 7th Australasian Remote Sensing Conference*, Vol. 1, Remote Sensing and Photogrammetry Association Australia, Floreat, Western Australia, pp. 51-60.

Robertson, P. K. (1990) A Methodology for Scientific Data Visualisation: Choosing Representations Based on a Natural Scene Paradigm. *Proceedings, IEEE Conference on Visualisation '90*.

Rogowitz, B. E. and Treinish, L. A. (1993) An Architecture for Rule-Based Visualisation. *IEEE Visualisation*, pp. 236-243.

Senay, H. and Ignatius, E. (1991) Compositional Analysis and Synthesis of Scientific Data Visualisation Techniques. *In: Patrikalakis (Ed), Scientific Visualisation of Physical*

Phenomena, Springer-Verlag, Hong Kong.

Senay, H. and Ignatius, E. (1994) A knowledge-Based System for Visualisation Design. *IEEE Computer Graphics and Applications*, pp. 36-47.

Tang, Q. (1992) A Personal Visualisation System for Visual Analysis of Area-Based Spatial Data, *Proceedings*, *GIS/LIS' 92.* Vol. 2, American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, USA, pp. 767-776.

Warner, J. (1990) Visual Data Analysis Into the 90's: *Pixel*, Vol. 1(1), pp. 40-44.

Author Information

Mark Gahegan is a lecturer in the Department of GIS,, Curtin University, PO BOX U 1987, Perth 6001, Western Australia. He may be contacted by phone: +619 351 3309, fax: +619 351 2819 or email: mark@cs.curtin.edu.au.

Ian D. Bishop, Chris Karadaglis

COMBINING GIS BASED ENVIRONMENTAL MODELING AND VISUALIZATION: ANOTHER WINDOW ON TH E MODELING PROCESS

ABSTRACT

Geographic Information Systems (GIS) have become increasingly valuable tools in environmental impact modeling and have been used in the evaluation of development prop osals in many ways. These may relate to specific environmental processes or to more intangible effects such as aesthetic or recreational values. A hitherto quite distinct activity has been the development of GIS driven visualization including realistic vi sual simulation. Typically these processes have been undertaken on separate platforms, by different people and in non-interactive time frames. The increasing availability of very high performance graphics engines offers new opportunities for combining int eractive environmental modeling with interactive visualization within immersive decision support environments. This paper describes the development of such a system combining GIS based modeling of environmental impact with high performance visual simulati on in a multi-channel graphics environment. Specifically it reports work with the Silicon Graphics Reality Engine as a common platform for GIS based visual effects modeling and visualization. This combination of the analytical approach with more subjectiv e visual assessment offers the best of both worlds in terms of information transfer and decision reliability. Techniques for testing the effectiveness of decision support are introduced.

INTRODUCTION

Environ mental change comes in two forms - planned and unplanned. We make determinations as to the direction of urban development or forest management practice but we have less opportunity to prescribe the extent of die-back, fires or salinity effects. In these c ases we can hypothesise and model the spatial progress of environmental change. We can, in some cases, predict the effects of changing management policy using computational models.

Computational models are commonly used in forestry, hydr ology, air pollution research, urban development prediction, and control of fire, salinity and traffic. Those models which use spatial data as part of their input are increasingly being linked to geographic information systems (GIS) - see many examples in Goodchild et al (1993). The numerical output of computer models can be extremely voluminous. In the case of forest management, for example, management decisions can affect timber production, water catchment properties, recreational values, aesthetic valu es, energy usage or employment opportunities. Natural resources management typically requires prediction of environmental changes over long time periods. The changes are also typically dependent on management decisions and constraints applied to the resou rce continuously through the management period. The complexity of the decision environment is further compounded by distribution of management options over large areas which vary substantially in their local characteristics, their proximity to sensitive a reas, the available access or their public visibility.

It has been strongly argued (McCormick et al, 1987) that visualization technology, meaning computer generated graphics and audio, is of unparalleled importance in dealing with such o utput. Persuasive anecdotal evidence exists that visualization has made important scientific insights possible, and it is asserted that the more interactive a visualization is, and the more data it can unambiguously present at once, the more useful it may be.

To date visualization of natural resources data has very often involved high levels of conceptual abstraction, or separation of data into accessible sub-groups. A classic example is the abstraction inherent in map generation by a geographic information system. The GIS also separates data into themes representing different aspects of the environment. Abstraction is not inherently bad, but achieving less abstract mappings is desirable, because it is likely to yield more universally understandable visualizations (Haber & McNabb, 1990). One style of visualization which is inherently less abstract is called the "natural scene paradigm" (Robertson, 1991). This involves creating a visual simulation that is highly evocative of real views of real situations.

As decision-making become increasingly an exercise in public consultation and compromise, decision support requires that all aspects of a project be clearly understood by the public. Bishop (1994) argues that th e non-scientific audience for computer generated information want abstraction minimised, information content maximised and the whole package digestible and non-threatening. This suggests the use of a visual realism approach which shows the information con sumer (decision-maker or public) what will/may happen under a variety of conditions and permit the consumer to explore the alternative environments using their natural sensory perceptions. Those not trained in interpretation of abstract information should be able to work at a more intuitive level. This calls for an interactive approach to the visualization process.

As we are increasingly inclined to undertake spatial and dynamic process modeling with a GIS as the underlying source of mode l data, it seems natural that the GIS should also be the source of data for visual simulation.

This suggests that decision support be considered as a two *part* process. It cannot be described as a two *stage* process because i deally the two components - modeling and visualization - operate together in an interactive computational environment. The modeling feeds the visualization, which influences the human operator who changes the modeling parameters (Figure 1).

The objective in linking the modeling process to one or more visualization procedures is achievement of high levels of information presentation together with high levels of interactivity showing change of time and the results of management decisions. U ntil very recently it was not possible to achieve simultaneous success with both these objectives. One had to choose between information content and interactivity. More powerful workstations, faster graphics and a more systematic approach to the process o f visualization are bringing the goal much nearer. However in the example described here the visualization is interactive but the modeling is not. The model - which is written within a GIS and which defines the visual impact of a highway proposal - is run , for a set of decision options, prior to initiation of the visualization. In the longer term, a more direct link between the modeling and visualization, such that adjustment of controls on the visualization screen would activate a new model run which in turn would generate new imagery, is anticipated.

A parallel study (Bishop & Karadaglis, 1996) concentrated on visual realism in the visualisation process. This study uses more abstract means to present the output of the modeling proce ss.

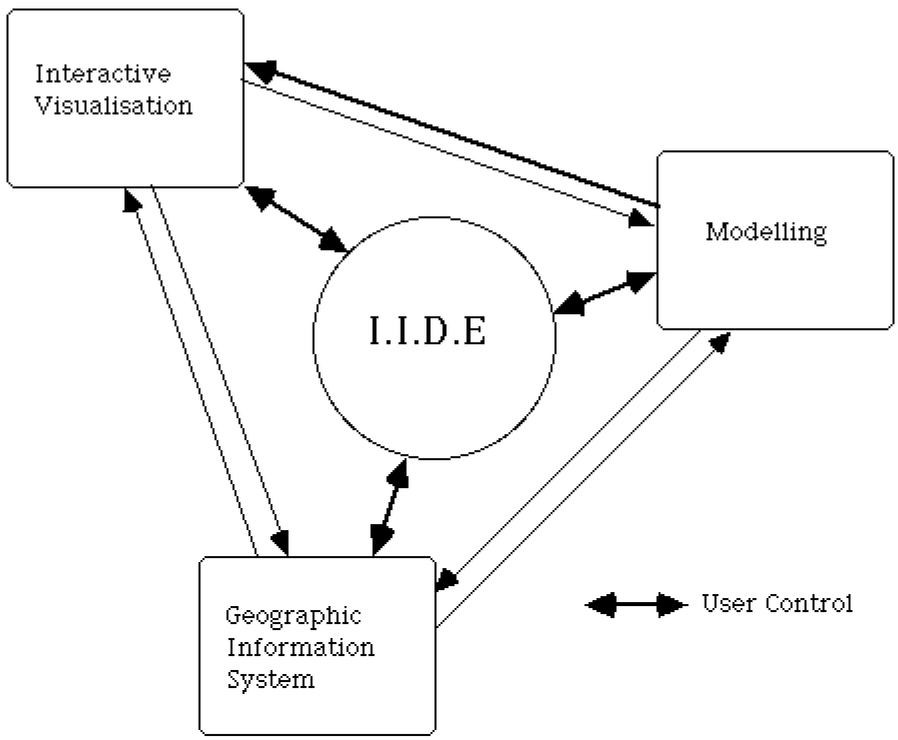


Figure 1. In an fully interactive immersive system all linkages are bi-directional. The key linkage being developed in this work is the feed from modeling to visualization.

< p>

VISUAL EFFECTS MODEL

The model has variables associated with each of three entities: the observer; the introduced object and the environment (Figure 2).

The Observer

The observer is any person with a line of sight to any object and who is potentially affected by the its visibility. It is assumed that the impact of the structures on any location depends upon both the expectations of each o bserver and the number of observers. In the model, existing roads can be used to give an approximation of the frequency of use. Estimates would be improved by traffic counts. Estimated scenic beauty - generated from an additional GIS based model (Bishop, in press; Bishop & Hulse, 1994) - can be used as a surrogate for probable user sensitivity on the assumption that the more attractive an area the more sensitive the user is likely to be to visual intrusion. From these two mappings cell sensitivity va lues are generated and normalised for the area.

The Object

Although the model will work with a single object most of the development has been based on analysis of linear features. Two special cases ar e electricity transmission lines and highways.

A transmission line has four major visual elements: the easement, the towers, the conducting wires and access tracks (Hadrian et al, 1988). Of these the towers are usually the most prominent feature. A transmission structure can be characterised according to its nominated location, a specific height and one of a restricted range of shapes. In the case of elements such as roads or pipelines there is no discrete element about which to compute i mpact values (except perhaps in the case of major interchange structures), the model therefore uses points along the feature - either turning points or points at regular intervals.

The visual effect of an introduced element is assumed to be highest when the observer is very close to the object and has a direct line of sight. The visual effect then decreases with distance. The current implementation of the model uses a linear distance-impact decay function although there is evidence (Hull & Bishop, 1988) that the impact is proportional to (1/distance).

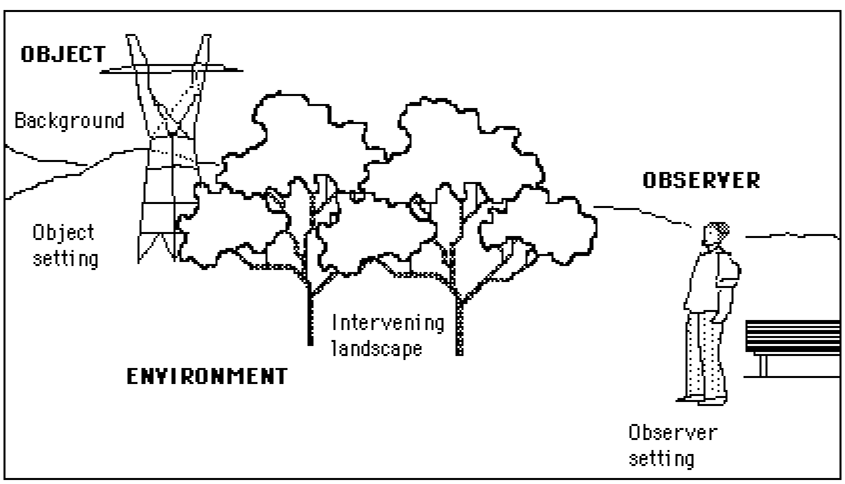


Figure 2. Major components of the model (after Hadrian et al, 1988)

The Environment

The environment is considered as a combination of landform and land cover. Topography and vegetation are the particular surface features included in the visual analysis.

The nature of the environment determines whether line of sight exists between the object and the observer (viewability), and whether the landscape between the object and the observer provides any significant visual absorption of the object. Vi ewability is derived from surface of the land, the height of surface features and the degree of visual penetration they permit. Different vegetation types permit different amounts of visual penetration. Height is kept as a grid value while screening perce ntage can be set interactively to conform to the vegetation types and patterns of the study environment.

In addition to providing screening of the object, the landscape between the object and the observer has a capacity to absorb the obje ct and still maintain its inherent visual character and quality (Yeomans, 1979). Landscapes, such as open water, which have very limited absorption capability have a default value of 1. More complex landscapes are assigned higher values by the user. Event ually a knowledge-based system will provide assistance to the user in setting parameters such as screening percentages and visual absorption ratios.

GIS IMPLEMENTATION

Modeling based on spatial data within a GIS can be undertaken as a stand-alone operation with data being transferred between the GIS and the modeling software via import and export facilities. This approach has the advantage that the modeling process itself may be very fast and efficient running in compiled and optimised code. While the modeling itself may be very fast the process of data transfer between the GIS and the model may be slow and awkward. The alternative chosen (Bishop & Robey, 1994) was for the m odel to run within the GIS package being written as a series of GIS commands: in the language of ARC/INFO as an AML (Arc Macro Language). Chou (1995) and Abel (1994) provide a more complete classification and comparison of the procedures for linkage of GI S with environmental modeling.

The process was divided into a separate modules for:

* object specification

* model operation

* analysis and mapping of results.

The original procedures for analysis and mapping (Bishop & Robey, 1994) are not described here because they are being replaced by the developing visualization system.

Object Specification

In the case of a transmission line the indivi dual tower positions and heights are kept in a file. The user may specify that some or all of the towers are to be included in the modeling. A choice can also be made whether to generate individual impact maps or a composite impact assessment.

Highway impact analysis is based on a road - or network of roads - defined as lines in an ARC/INFO coverage. The user chooses to base the analysis on either the nodes and vertices defining the roads - a quick but not very accurate approach - or on all the points along a road as derived by first doing a LINEGRID conversion followed by a GRIDLINE conversion. This creates a coverage in which there are points all along the road at intervals roughly equal to the grid size. Analysis then takes considerab ly longer but is also more complete.

Model Operation

Based on uniform reduction with distance, the visual effect is initially radially symmetrical about a tower or highway point. The COSTDISTAN CE function is used to determine the distance over which the effect of a single structure is felt to be significant. However, the intervening landscape's capacity to screen and to absorb visual effects results in different observers, at equal distances fr om a tower, being affected differently. Thus, a grid of visual absorption is incorporated into the COSTDISTANCE function as the 'cost-grid' to act as an impedance to the visual impact.

Once the initial distance-decay associated with visua l impact has been modelled a viewshed analysis (VISIBILITY) is done determining line of sight between a point midway up the tower and each cell falling within a specified distance range of the tower, taking into account screening height and percentage.

The impact of a line is taken as the sum of the impacts of the individual towers - or the sample points along the road in the case of a highway (Figure 3).

THE VISUALIZATION TOOLKIT

The visua lization component of the work is based on IRIS Performer Toolkit (Rolf & Helman, 1994) running on an SGI Crimson based Reality Engine. Performer provides a host of C and C++ based commands which can be called from user developed programs. These comma nds provide for the loading of various data formats, interactive movement of the view position and of objects within the modelled environment and development of GUI 'widgets' for control of other aspects of the visual presentation. The loaded data is spat ially segmented to provide for optimum redraw speed making maximum use of the Reality Engine graphics pipeline. The Reality Engine provides for hardware based polygon texture mapping with no time penalty relative to simple surface shading. Performer, with Reality Engine, has been used in the urban planning context (Liggett & Jepson, 1995) and in battlefield simulations but we are not aware of its previous application to regional or natural resources planning.

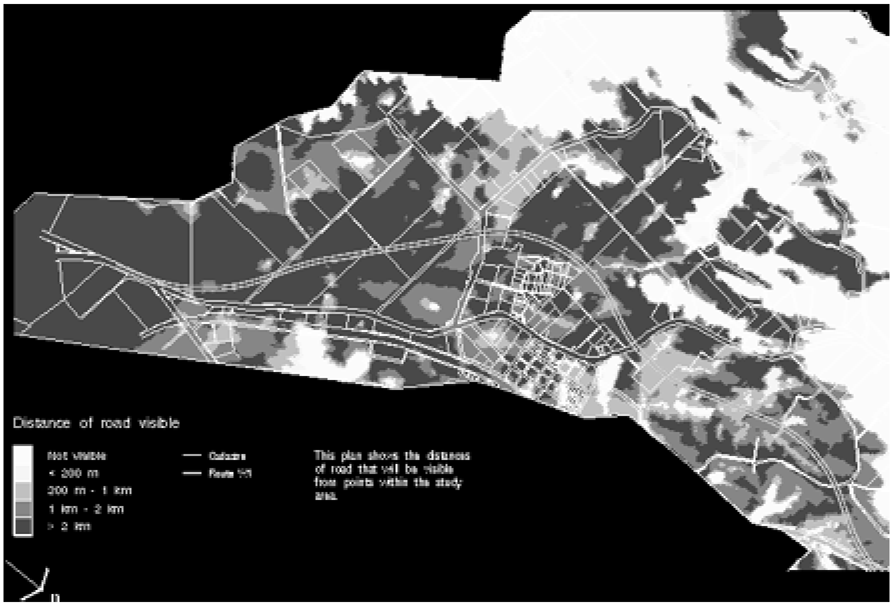


Figure 3. Conventional GIS generated map output from visual impact model

DEVELOPMENT OF INTEGRATED SYSTEM

The conceptual ideal is for three independent systems to be closely coupled thro ugh transparent interfaces and communication protocols: transformers, constructors and accessors (Abel et al, 1994). In practice, however, compromises to this ideal were necessary. The comparative slowness of the modeling process, the proprietary nature o f the GIS data structures and the flexibility provided by working in C in conjunction with the Performer toolkit mean that the development work described here is essentially focussed on the visualization component. The model outputs are written into the G IS as attributes to the analytical polygons and the GIS then exports these results and the height field in raster format.

The visualization procedures described and illustrated in this paper were designed initially to work in a forest ma nagement context (Bishop & Karadaglis, 1996). In that instance the visualization sought to represent in a visually realistic way changes in forest conditions over time. In adapting those procedures to work with visual effects data and to represent th e output in a more abstract manner we are to some degree under-using the capabilities of the system.

Visualization Data

The data to be visualised consist of the following :

* An ARC/INFO fi le containing sampled elevation data for the Black Forest Section of the Calder Highway, Victoria realignment. The file also contains the number of rows and columns, as well as the extent of the data. The 3 dimensional DTM is constructed using this data.

* The results of simple viewshed and more complex visual impact modeling in ARC/INFO format, sampled at the same resolution as the DTM. These files are referred to as visual effects files (VEFs). One VEF exists for each sample point of each rout e under each set of modeling assumptions (i.e. each model run).

* Definitions of the various variables which correspond to the decision variables of the model runs. The variable information is defined in a Variable Definition File (VDF).

* Definitions of the various thumbnail textures (or colors in this example) which are used to change the appearance of the original DTM model. An internal variable ('vis_imp') is calculated for each cell based on the VEFs, and the value of this variable d etermines which color or texture is applied in that cell.

* An overall - existing conditions - aerial image of the terrain. This texture which is mapped onto the DTM surface could be a satellite image or aerial photograph or some mapped informati on (e.g. the cadastral boundaries or the stream network). This information blends with the colors representing visual impacts. The user may also switch between an aerial photograph and mapped information or switch off this texture altogether.

Rules for visualising the data :

Model runs can be the combination of several user defined variables (specified in the VDF). For each permutation of the values of the user defined variables, there exi sts a model run, and a corresponding set of VEFs. The set of VEFs represents the results of visual impact studies for all points along the route.

The initialisation process involves reading in all VEFs. Specifying the large number of VEF filenames is not necessary due to the VDF loader, which uses data from this file to automatically create a set of filenames to load. In addition, the VDF contains information which builds GUI widgets (e.g. defining maximum impact distances or rates of vis ual absorption) which can be used to select the particular model to view while the simulation is running.

Once all VEF data is loaded, the simulation begins. Two GUI sliders, 'Route Start Point' and 'Route End Point' are used to specify t he segment of the route. Once a segment is defined, the cumulative visual impact effects for each point in the segment are calculated, and colors or thumbnails are assigned to each cell. These cells are then rendered and the simulation continues.

Program description :

The application program, "`ide`", was built on top of an IRIS Performer demonstration program, "`perfly`". Customization to our application involves initialising and reading in al l the VEFs and building an appropriate scene graph, which is then rendered at a fixed frame rate.

The elevation ARC/INFO file (*.e00), the Thumbnail Definition File (TDF), and the Variable

Definition File (VDF) are specified through the c ommand line as follows:

```
ide calder.e00 calder.tdf calder.vdf
```

The full structure of the TDF and VDF files is explained in Bishop & Karadaglis (1996).

These files are loaded during th e initialisation phase, their cumulative effect is computed and the impact on the landscape at each grid cell is calculated. If found to cause mappable visual impact, an appropriate color or texture thumbnail is selected and rendered at that position.

Program Initialisation :

The initialisation sequence includes:

***** Load the VDF and TDF files, and build a list of VEF filenames.

***** Load the elevation ARC/INFO file, and bui ld the default DTM.

***** Load all VEF ARC/INFO files, and calculate the cumulative effects of the highway for each cell for each variable value permutation.

***** Examine the TDF and map the colors or thumbnail textures to the appr opriate cells which are visually affected by the highway.

***** Using vertex and surface normals from the default DTM, and texture scaling coordinates from the TDF file, build a DTM model for each variable value permutation.

***** At tach each visual effects model to a switch node, using an internal index value which can later be used to select the required DTM to be rendered.

* Initialise the environment model. This includes defining the sun as an infinite light source, and s etting the horizon colours.

***** Initialise and build the GUI interface, including any new variable widgets that are defined in the VDF.

Real time visualization :

Once initialised, the "`ide`" application enters the real time component. The default DTM is displayed. Motion is accomplished by choosing a motion model and using the mouse. The following motion models (from "`perfly`") are currently available :

* "fly" : Allows changes in heading and pitch using the mouse. Acceleration is possible using the mouse buttons. No changes in roll are allowed.

* "drive" : uses terrain following to determine the pitch and roll relative to the direction o f the

surface immediately below the viewpoint. A constant height above the terrain is maintained. Heading and acceleration changes are accomplished using the mouse.

* "track" : permits no viewpoint motion but allows the user to change the positio n and orientation of the DTM in all 3 axis.

The viewpoint can be repositioned at any time to just outside the bounding sphere of the DTM, or to its centre, using GUI buttons.

Selecting a visual impact result to display is accomplished through either using the menu for user defined variable ("distance" is specified by start point and end point sliders, all others are implemented as menus), or by using a keyboard shortcut. Once a user defined variable widget has been sele cted, an internal index is used to switch to the correct visual effects map for visualization.

Zoom is possible by changing the Field of View slider in the GUI. The DTM can be scaled in the Z direction to accentuate changes in the shape of the terrain. Surface normals are currently not updated in this mode.

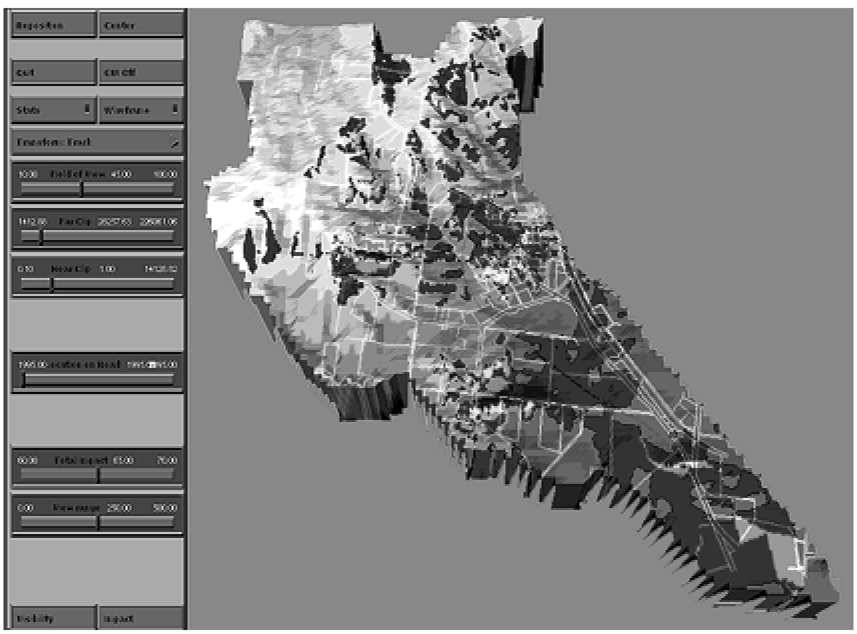


Figure 4. View of model output on three dimensional model of area. The mouse moves the 3D model in real time for rotation or fly-overs. The controls at the left allow the user to select the road segment to be impact mapped or to change other modeling parameters (e.g. assumptions about maximum impact distances or rates of visual absorption).

CONC LUSION

There are a number of readily accessible enhancements to the visualization system developed to date. These include:

* Add Level of Detail switching (a performer node which selects between high resoluti on and low resolution models of the same region based on the distance of the viewpoint from the centre of the model). This will improve the rendering performance of the application since it will allow larger DTMs with effectively the same resolution to be rendered.

* Add window plot interfaces to display the status of additional user definable variables, such as total visual impact. These windows may be implemented using overlays, or by using custom textured polygons (billboards) which ar e moved to ensure they remain visible regardless of the viewpoint position.

* Allow for subsampling of the ARC/INFO files, to either reduce the total number of polygons, or to create a lower resolution scene graph for the collision detect ion library (to improve collision performance).

* Add the ability to sequence through the DTMs rather than switch using the GUI or keyboard. This could be used to animate the visual impact through the construction period for one particula r model, even while the viewpoint is in motion. The application to other modeling projects that project change over time is even more apparent.

* Add stereo graphics to a create more immersive decision environment (Ribarsky et al, 1994). Head-mounted displays and motion tracking provide the opportunity to create a decision environment in virtual reality. A single decision-maker may move through the virtual environment through head or hand movements, while controlling the time or the decis ion variables using a virtual interface. Using multi-channel output several protagonists could be immersed simultaneously ensuring that each is seeing the same visualised effects of changes in model parameters arrived at collectively.

* C reate new motion models which can be defined as splines through user selectable points. Animate the DTMs through time steps while moving.

* Add stress management by building lower resolution DTMs and configuring Performer to drop to a low er resolution model if the graphics load exceeds a particular value. This is particularly important for fixed frame rate, immersive applications (to minimise simulation discomfort).

When the programs which comprise the system have been d eveloped further it will then be appropriate to undertake some rigorous testing of the visualization platform. It has been implicit in this work that enhanced means of presentation of the output of GIS based modeling will enhance understanding of the resu lts and lead to better environmental decision making. This assumption needs to be tested and the value of development in this direction demonstrated. Two approaches to systems evaluation are foreshadowed: information transfer and decision convergence.

In each application specific context there will be some potential users who have an advanced knowledge of the applications domain and others who do not. We plan to test both expert and non-expert decision-makers through exposure to the visu alization environment. During this exposure they will undertake a series of tasks to determine the accuracy of their assessment of complex relationships. Subjects will be asked to estimate parameter values at indicated locations within the data, and to co mpare values at two or more different locations. The accuracy and variation of these responses will be compared with actual values, as a function of the visualization parameters. We will have both "look here" situations, where users areas are indicated b y a marker in the display, and "look where" situations, where users are asked to find the location by description.

We will also test directly the decision support capacity of the systems. This will be achieved using a small group of exper ts within the particular applications area and determining their level of agreement on some planning or management decision based on the different visualization paradigms. The experts might be asked, for example, to select the route for a highway bypass a round a town. If given no information at all a high variance in the position of the chosen routes can be anticipated. As the level of available information - including social, economic, engineering and aesthetic model outputs and visual simulations of dri ving along the routes - increases the chosen routes should converge and the variance decrease.

Our analysis will also make use of measurable aspects of task completion such as overall time, or recurrence of certain actions. The interacti ve patterns of users with the modeling/visualization systems will be monitored by the computer itself and recorded in explicit log files. At the same time the users actions and reactions will also be recorded on video-tape to support the analysis. The res ults of the information transfer and decision agreement experiments will provide a clear indication of the value of the visualization paradigms.

ACKNOWLEDGMENTS

This visualization development was funded by the Australian Research Council. The visual modeling process was developed as part of a larger body of research supported by VicRoads. The data sets were provided by VicRoads in conjunction with a visual impact study undertaken by the Gerner Consulting Gr oup.

REFERENCES

Abel, D. J., Kilby, P. J., & Davis, J. R. (1994). The systems integragion problem. International Journal of GIS, 8, N1, 1-12.

Bishop, I. D., & Karadagl is, C. (1996). Use of interactive immersive visualization techniques for natural resources management. In Visual Data Exploration and Analysis III (Electronic Imaging '96), . San José, CA: The Society for Imaging Science and Technology.

Bishop, I. D. (1994). The role of visual realism in communicating and understanding spatial change and process. In D. Unwin & H. Hearnshaw (Eds.), Visualization in Geographic Information Systems London: Bellhaven Press.

Bishop, I. D. (in press). Comparing regression and neural net based approaches to modeling of scenic beauty. Landscape and Urban Planning.

Bishop, I. D., & Flaherty, E. (1990). Using video imagery as texture maps for model driven visual simulation . In Proceeding of Resource Technology '90, (pp. 58-67). Washington, D.C.

Bishop, I. D., & Hull, R. B. IV (1991). Integrating technologies for visual resource management. Journal of Environmental Management, 32, 295-312.

Bishop, I. D., & Hulse, D. W. (1994). Predicting scenic beauty using mapping data and geographic information systems. Landscape and Urban Planning, 30(1-2), 59-70.

Bishop, I. D., & Robey, M. (1994). Implementing an e nvironmental impact model within a geographic information system. In AURISA'94, (pp. 281-291). Sydney: Australasian Urban and Regional Information Systems Association.

Chou, H. C. (1995). Integrating Environmental Models and GIS. In Int ernational Symposium on Geographic Information System and Environmental Protection, . Taipei, Taiwan: Environment Protection Administration, Republic of China.

Goodchild, M. F., Parks, B. O., & Steyaert, L. T. (Eds.). (1993). Environme ntal Modeling with GIS. Oxford: Oxford University Press.

Haber, R.B. and McNabb, D.A. (1990) Visualization Idioms: A Conceptual Model for Scientific Visualization Systems, in Visualization in Scientific Computing, Nielson, G.M. and Sh river, B., editors, IEEE Computer Society Press, Los Alamitos, Calif.

Hadrian, D. R., Bishop, I. D., & Mitcheltree, R. (1988). Automated mapping of visual impacts in utility corridors. Landscape and Urban Planning, 16, 261-283.

Hull, R. B. I., & Bishop, I. D. (1988). Scenic impacts of elctricity transmission towers: the influence of landscape type and ovserver distance. Journal of Environmental Management, 27, 99-108.

Liggett, R.S. and Jepson, W. (1995) Use of real time visual simulation technology for urban planning/design decision making, Proceedings 4th International Conference on Computers in Urban Planning and Urban Management (v1, 51-64), Melbourne.

McCormick, B.H., DeFanti T .A. and Brown M.D. (eds) (1987) Visualization in scientific computing. Computer Graphics 21(6), ACM SIGGRAPH.

Perkins, N.H. (1992) "Three questions on the use of photo-realisitic simulations as real world surrogates", *Landscape and Urban Pla nning*, 21, 265-267.

Ribarsky, W., Bolter, J., Op den Bosch, A., & van Teylingen, R. (1994). Visualization and Analysis Using Virtual Reality. IEEE Computer Graphics and Applications(January), 10-12.

Robertson P.K. (1991) A m ethodology for choosing data representations, *IEEE Computer Graphics and Applications*, May, 56-67.

Rolf, J., & Helman, J. (1994). IRIS Performer: a high performance multiprocessing toolkit for real-time 3D graphics. In SIGGRAPH 94, (pp. 381-394). Orlando, FL: ACM SIGGRAPH.

Yeomans, W. C. (1986). Visual impact assessment: changes in natural and rural environment. New York: John Wiley.

AUTHOR INFORMATION

Ian D. Bishop Chris Karadaglis

Director Research Staff

Centre for Geographic Information Systems and Modeling

The University of Melbourne

Parkville, V 3052, AUSTRALIA

Phone 61 3 9344 6944

Fax 61 3 9347 2916

Ian_Bishop@mac.unimelb.edu.au

James Kuiper, A. Ayers, R. Johnson, M. Tolbert-Smith

Efficient Data Exchange: Integrating a Vector GIS with an Object-Oriented, 3-D Visualization System

Presented at the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, USA, January 21-25, 1996

Sponsored by the National Center for Geographic Information and Analysis (NCGIA)

ABSTRACT

A common problem encountered in Geographic Information System (GIS) modeling is the exchange of data between different software packages to best utilize the unique features of each package. This paper describes a project to integrate two systems through efficient data exchange. The first is a widely used GIS based on a relational data model. This system has a broad set of data input, processing, and output capabilities, but lacks three-dimensional (3-D) visualization and certain modeling functions. The second system is a specialized object-oriented package designed for 3-D visualization and modeling. Although this second system is useful for subsurface modeling and hazardous waste site characterization, it does not provide many of the capabilities of a complete GIS. The system-integration project resulted in an easy-to-use program to transfer information between the systems, making many of the more complex conversion issues transparent to the user. The strengths of both systems are accessible, allowing the scientist more time to focus on analysis. This paper details the capabilities of the two systems, explains the technical issues associated with data exchange and how they were solved, and outlines an example analysis project that used the integrated systems.

INTRODUCTION

Scientists in the Environmental Assessment Division (EAD) of Argonne National Laboratory are involved in a wide variety of projects and use many different software packages to conduct research. Although the variety of software packages used expands their capabilities to analyze and model information, it also can result in bottlenecks when data cannot be easily transferred from one software package to another. This situation is neither new nor unusual. Both individual software developers (at the application level) and organizations for standards development (at the industry level) are addressing this problem.

Two software packages frequently used by EAD are ARC/INFO(TM) (Environmental Systems Research Institute, ESRI Inc.), a Geographic Information System (GIS); and SiteView(TM) (ConSolve Inc.), a three-dimensional (3-D) visualization and analysis tool. Although both are designed for spatial data analysis with overlapping functionality, they each have unique capabilities. In the case study of a Department of Energy (DOE) project presented here, it is clear that both systems provided essential capabilities not available in the other. The case study involved a site at a DOE facility that was historically used to dispose of hazardous waste.

The two systems had only one common exchange format for unattributed line and polygon data: the AutoCAD(TM) (Autodesk Inc.) Drawing Exchange Format (DXF) (Autodesk 1993). Exchange of other data is limited to column-formatted text files. As shown by the case study, efficient data exchange allowed these two tools to be integrated for the project, taking advantage of the unique capabilities of each.

BACKGROUND

ARC/INFO is an extensive GIS with vector and raster capabilities including import, georeferencing, editing, analysis, and output. The system provides the capability to generate 3-D surface views, but it is not designed to model or display subsurface information. For the DOE case study, general geographic data for base maps of the study were available in ARC/INFO. The raster analysis module was used for non-intrusive geophysical survey data, including input, georeferencing, interpolation, and processing to produce imagery. It was also used to input and georeference scanned aerial photographs of the study area.

SiteView is more limited and lacks several of the major components of a GIS, but it is specifically designed for 3-D visualization and analysis of surface and subsurface site characterization data. It can easily import tabular data and point graphics, but for more complex spatial features it is limited to the DXF file format or manual input by the user. Data imported with the DXF format is for display only, and cannot be attributed beyond its graphics-related information. SiteView has an easy-to-use graphical user interface, and is much less expensive than ARC/INFO. In the DOE case study, SiteView was used to store and visualize subsurface data.

The difference in data models was the most significant obstacle to be overcome in integrating the two systems. ARC/INFO has a topological, feature-based structure for graphics and a relational database for tabular attributes. A linear item, such as a road, is given a line feature type. Attributes describing roads are stored in a relational database and attached to the lines by a system-maintained link. The line feature type maintains topology so line connections are known and can be used for analysis. Land-use areas would be stored with a polygon feature type in a separate data set. Polygon topology allows adjacent polygons and their shared boundary to be identified. In general, the feature type structure is designed for analysis, but relationships between data sets, such as selecting roads within a land-use polygon, must be calculated by the system or determined by the user in some manner. Although feature types such as points, lines, or polygons are designed into the system, the data set name and design of tabular attribute fields are user defined.

In contrast, SiteView uses an object-oriented data model to store information. The set of objects, such as well points or soil polygons is pre-defined, cannot easily be modified by the user, and is designed specifically for site characterization data. Land-use areas in this data model are stored as land-use objects containing both polygon coordinates and tabular attributes. The available attributes are pre-determined by the object's structure. Data for sampling wells are stored as well objects with both coordinates and attributes. Data for samples from the wells are stored as sample objects. The set of objects can be organized and used in more complex ways than a relational data model easily permits. Samples, when linked to a well, inherit the geographic location of the well. A set of wells can be associated with a land-use area. Once this is done, the set of samples can be identified based on land use using links inherent in the data. SiteView does not maintain topological relationships. For example, no function exists to determine if two polygons are adjacent. The line between them is duplicated rather than shared and may not necessarily be coincident.

Both systems and data models provided useful and unique capabilities, but the significant differences and lack of an efficient method of data exchange limited their combined use.

IMPLEMENTATION

A project to improve data exchange between the two systems was funded by the DOE. The work was designed to focus on some immediate needs of DOE environmental restoration projects, to provide generally applicable extensions to SiteView, and to result in code that could be useful for other similar data transfer applications.

The level of integration was limited to data translation using an intermediate data file, primarily because the systems differ significantly in data models as well as representation of more complex graphic elements, such as polygons. The method of integration also addressed proprietary issues associated with both software packages and transfer of data between separate users or systems.

Design of the Intermediate File Format

In early stages of the project, many established file formats were examined. In particular, a design based on the Spatial Data Transfer Standard (SDTS/FIPS-173) (NIST 1992) was considered. Import and export of SDTS files was newly supported by ARC/INFO (ESRI 1995a) and the standard provided detailed guidance for file formatting and other issues. SDTS was designed to be an all-encompassing standard and had many elements beyond the scope or needs of the project. Although it was an excellent source for design concepts, it did not offer an ideal solution.

ArcView(TM) (ESRI, Inc.) Shapefiles were also considered. Import and export of this format was also supported in ARC/INFO. Shapefiles are binary and store graphics and attributes in a nontopological format (ESRI 1995b). This is an example of a software development organization designing a format optimized for its product. SiteView lacks a similar transfer format geared to its data model. It would be useful for SiteView to support the Shapefile format, but a format optimized for SiteView would provide more options for data exchange from a variety of sources.

The approach selected was to design and implement an ASCII file format that was uniquely adapted to SiteView and its data model. This approach allowed the format to be easily adapted to other uses, since ASCII files can be easily created or modified by users. The ASCII file consists of an informational header, a structure definition, and data records (Figure 1). The structure definition identifies the object type, geometric type, and the data attributes. The data section begins by identifying the object type and the data elements follow. Each data element begins with a tab separated list of attribute values followed by one or more coordinates, depending on the object type. Multiple structure definition and data sections are permitted in the file. A detailed file specification was documented by ConSolve (ConSolve 1995).

```
***SLF_FILE***
/work/doe/translate/roads83.slf
Mon Dec  4 08:50:35 1995
SiteLink v1.0
This data set contains road lines for the DOE study area.
;;;
Road
Polyline
Name|30|String|-1
PavementType|10|String|-1
;;
;;;
Road|93
    1|paved|
 3.95822218671853e+05| 7.80876320926992e+04| 0.00000000000000e+00
 3.95823002518686e+05| 7.80952362633142e+04| 0.00000000000000e+00
 3.95823860547643e+05| 7.81000000000000e+04| 0.00000000000000e+00
;
   93|dirt|
 3.95606244063250e+05| 7.81000000000000e+04| 0.00000000000000e+00
 3.95612311693291e+05| 7.80878534624764e+04| 0.00000000000000e+00
 3.95634650550764e+05| 7.80479487511311e+04| 0.00000000000000e+00
;
;;
;;;
```

Figure 1. Example of the transfer file format. Tab characters are shown as "|" characters for legibility. Only the first and last data records are shown.

Approach to Coding

The programmers had several choices for coding the ARC/INFO translator. One option was to use Arc Macro Language (AML) to transfer the information between ARC/INFO and the file. This method had the advantage of being compatible with all ARC/INFO-supported platforms, but the disadvantages of slow execution speed and limited language capabilities.

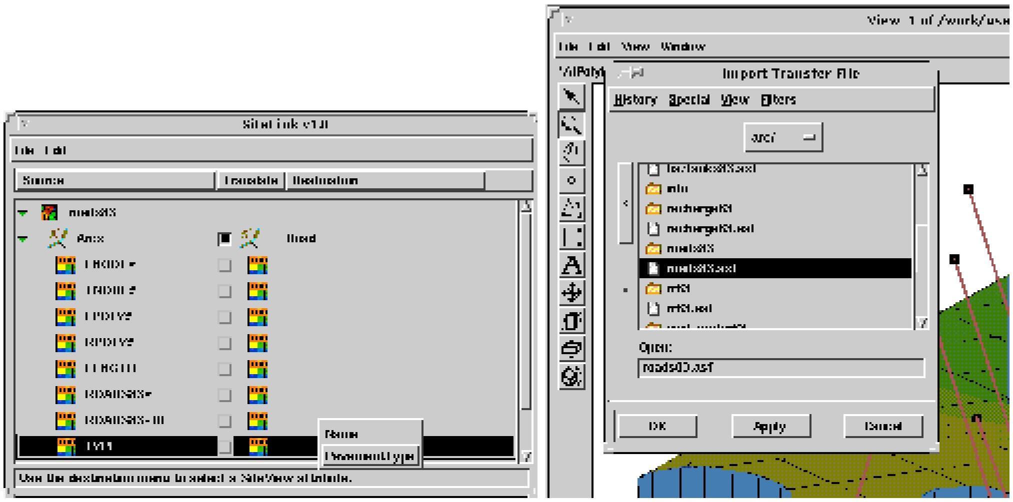
The coding approach chosen was to use ARC/INFO software development libraries (ArcSDL(TM)) and C code. The object libraries allow FORTRAN or C code to use most of the low-level routines in ARC/INFO to access data in its internal format. The final result can be a standalone application. The main disadvantages with this approach are portability (because the libraries and compiler are sold separately for different platforms), and the requirement of an ARC/INFO license to run the program. The positive factors that led to choosing this option were the speed improvements of compiled code, the ability to do all processing within one program, direct access to all ARC/INFO internal data types, and more flexibility in the design process. Also, the code developed would be useful for other future integration projects. The Galaxy Application Environment(TM) (Visix Software, Inc.) was used for the interface. It is a cross-platform development tool with an efficient graphical user interface (GUI) builder. It is also used by ConSolve to develop their products.

The User Interface

Import of the transfer file was added to the SiteView "File" menu as a new option. On the ARC/INFO side of the translation, the program was coded as a standalone application, SiteLink, with a graphical user interface. Most of the translation work is done in this application.

One of the more difficult issues was how to efficiently map data sets and attribute data fields from ARC/INFO's user-defined approach to SiteView's predefined object classes. The method developed was to add the SiteView object dictionary as a resource to the SiteLink translator. This allowed ARC/INFO data to be mapped to SiteView objects at the time of export. For example, the process to create a transfer file for an ARC/INFO roads data set and import it to SiteView would be as follows:

- In SiteLink, use the file browser to load the ARC/INFO roads coverage.
- Open up the list of arc attribute table fields.
- Choose a destination SiteView object class. The interface limits the user to object classes that can contain lines, in this case, the "Road" object type.
- Choose SiteView attribute slots for the ARC/INFO attributes to be transferred. The user is limited to unused SiteView attribute slots with a proper data type. Here, the INFO item "TYPE" is mapped to the SiteView attribute slot "PavementType" (Figure 2a).
- Save the data as a transfer file using the file menu.
- In SiteView, use the "Import Transfer File" option in the "File" menu to load the contents of the file (Figure 2b).



(a) (b)

Figure 2. (a) Using the SiteLink translator to translate an ARC/INFO roads data set. The "Road" object type has been assigned, and the PavementType attribute slot is being assigned for the INFO attribute "TYPE." (b) Importing the roads SiteLink transfer file into SiteView.

Solutions for System Differences and Limitations

The translation program needed to provide useful functionality and appropriately preserve significant information in the data. This required deciding which data types to support and designing algorithms to overcome differences in geometric object representation. Not all ARC/INFO data types were included for translation. SiteView does not support raster data storage, so the ARC/INFO Grid format was excluded. 3-D surfaces can be generated in either system, and were not included. Annotation and nonspatial attribute tables were omitted because they could be transferred by existing means.

Differing polygon representations was the major geometric issue. ARC/INFO polygons have a list of lines that define a closed boundary. The attributes are linked to a point inside the polygon. Lines are shared between adjacent polygons and holes within polygons are recognized. In SiteView polygon objects, the label point is not necessary and the attributes are an integral part of the object. The full closed perimeter of each polygon is specified and lines between adjacent polygons are duplicated. Holes are not possible, and must be represented as additional polygons on top of another one. The SiteLink transfer file is modeled after the SiteView format. ARC/INFO polygon holes are treated as separate polygons.

Regions in ARC/INFO are composed of one or more areas enclosed by rings of arcs. The discontiguous areas are treated as one attributed unit, can contain holes, and can overlap with other regions. In SiteLink, regions are output as a set of polygons, each having the region attributes. Region holes are also output as polygons, but attributes are assigned as null values. The programming challenge was to separate holes from additional discontiguous areas, since this is only represented by the direction in which the vertices are stored. Clockwise storage indicates a discontiguous area and counter-clockwise storage indicates a hole. Direction is determined by calculating the area, assuming that vertices are stored in a clockwise direction. If the calculated area is negative, then the vertices are stored counter-clockwise and the polygon is a hole.

SiteView stores point data in three dimensional form, while ARC/INFO does not. In ARC/INFO, elevation is stored in an attribute field. The SiteLink translator allows SiteView Z-coordinate values to be populated from an ARC/INFO numeric attribute. This was implemented efficiently in the user interface by adding a Z field to the menu choices for mapping numeric items.

For the case where data need to be added to SiteView, but an appropriate existing object class did not exist, generic object classes were added. These classes, named after the type of feature, (point, line, or polygon) have a set of generic attribute fields. In these objects, the ARC/INFO data set name is automatically transferred to a SiteView attribute so that the data can be easily identified and grouped in SiteView.

CASE STUDY

The combined capabilities of the GIS system, translator, and 3-D visualization package were used to help study a DOE site with subsurface contamination. The site included an unknown number of waste disposal pits that had to be located and sampled. The site characterization objectives were to determine the number, locations, and sizes of the disposal pits, and to determine the nature and extent of any contamination emanating from the pits. The first phase of the work used non-intrusive geophysical techniques including magnetic gradiometry, electromagnetic surveying, and ground penetrating radar profiling to locate the pits. The second phase was an intrusive sampling program targeting soils beneath and adjacent to identified disposal pits.

ARC/INFO and SiteView software were used to develop most of the figures produced for the DOE work. In combination they provided the ability to input, process, analyze, model, and visualize a variety of spatial data derived from existing data sources and field data collection. With the SiteLink translator, data that resided only in ARC/INFO was efficiently translated to SiteView.

ARC/INFO provided support for data import, georeferencing, and surface data interpolation and display. A base map and 1962 aerial photograph of the study site is shown in Figure 3. Several of the main disposal pits are visible in the photograph, as well as smoke from open burning at a nearby conventional landfill. Figure 4 depicts the results of the magnetic gradiometer survey of the site. This survey method identifies variations in magnetic fields at the ground surface and the light and dark areas represent positive and negative magnetic fluctuations. It successfully identified the disposal pits.

SiteView was used for most of the intrusive survey data because of its ability to produce 3-D views of subsurface data, fence diagrams, and analysis of disposal pit volumes. In Figure 5a, a subsurface view of the study area is shown with the disposal pits and slant bores. The two disposal pits identified in Figure 4 are highlighted. On the surface, road lines and forest polygons translated from ARC/INFO are shown. A groundwater depth surface is shown in Figure 5b. This surface was created in SiteView using the wells and depth-to-groundwater data.

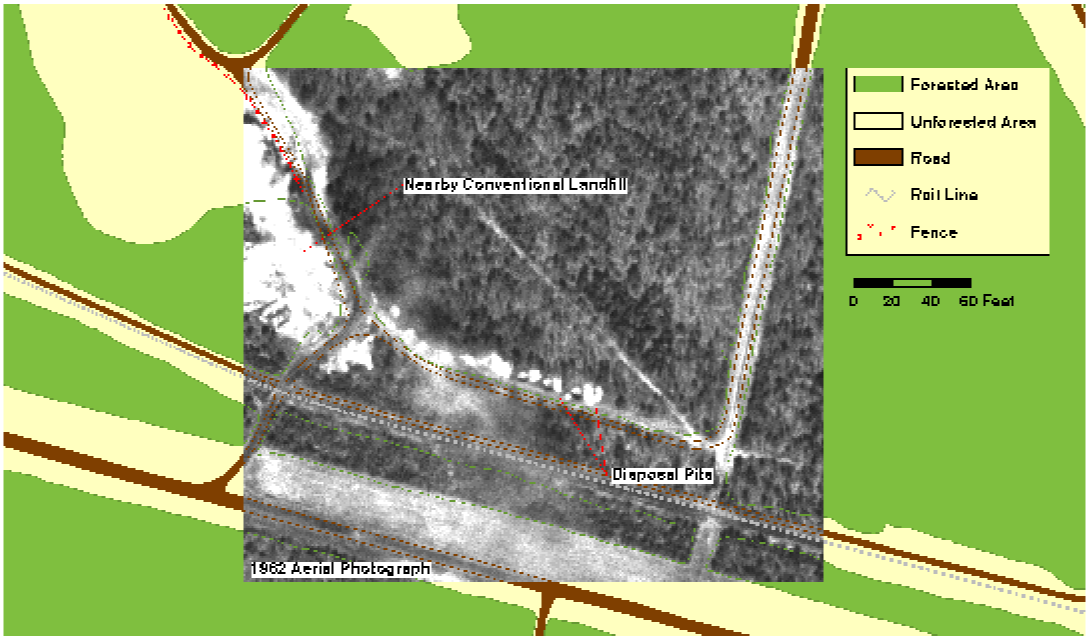


Figure 3. Base map produced in ARC/INFO with 1962 aerial photograph showing the study area.

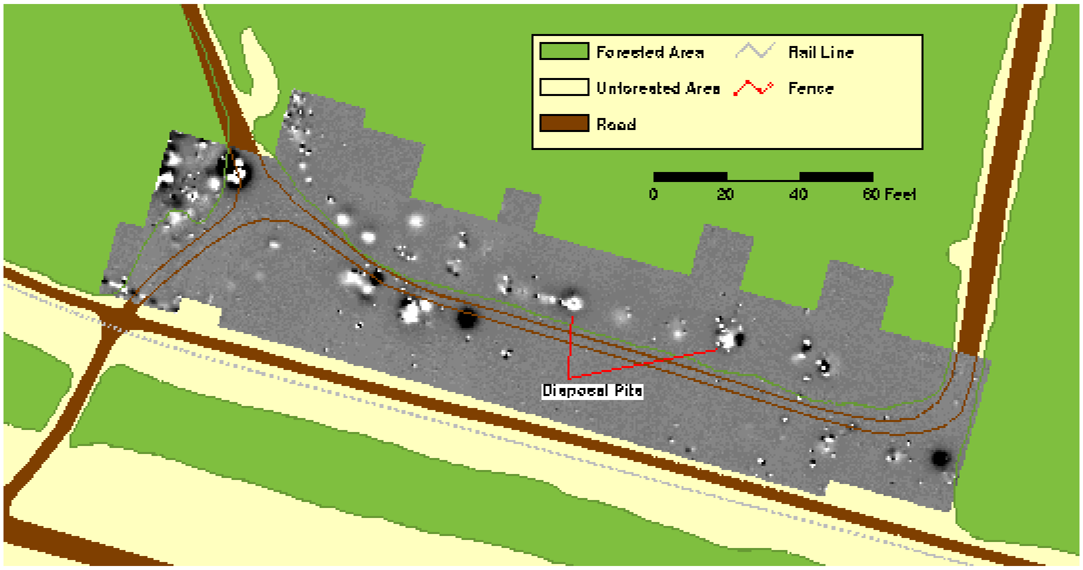
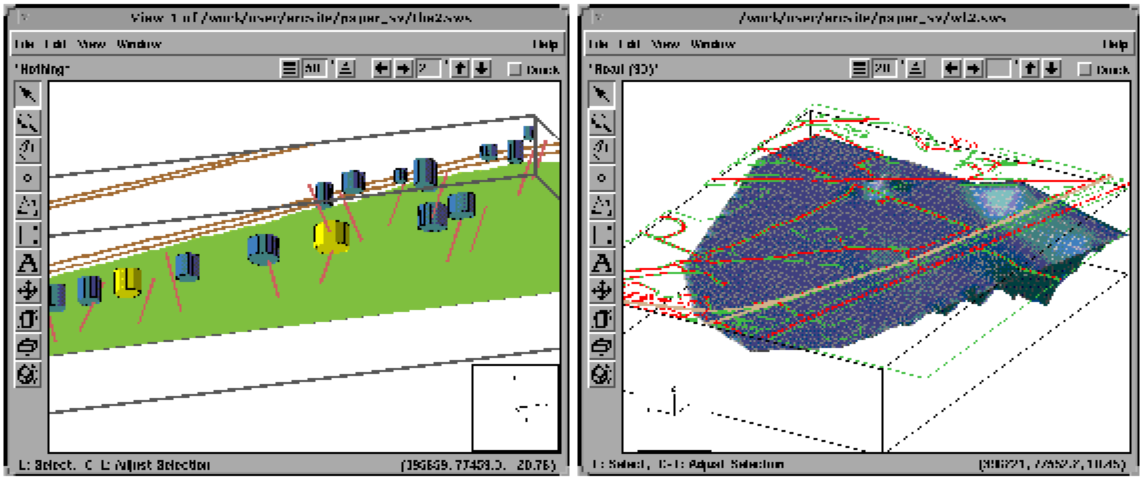


Figure 4. Nonintrusive magnetic gradiometer survey data shown in ARC/INFO.



(a) (b)

Figure 5. (a) Subsurface view of the study area in SiteView showing disposal pits and slant bores. Surface map layers were translated from ARC/INFO. (b) Groundwater depth visualized in SiteView with translated base map layers superimposed on the surface.

CONCLUSIONS

The SiteLink translator provides a simple method for importing data to SiteView and insulates the researcher from some of the more complex and detailed translation issues. This enables more productive analysis, as is evident in the case study. Much of the detail involved in reformatting information and mapping it to the target system is simplified for, or transparent to, the user.

The SiteLink translator benefits ARC/INFO users by giving them new 3-D visualization capabilities and SiteView users by giving them streamlined access to data from a widely used GIS. The transfer file format was designed as a general purpose file for SiteView and can be easily implemented for exchange with other systems. For the ARC/INFO user, SiteLink can also be implemented for general purpose file export, especially for transfer to object-oriented systems or those having similar data representation designs.

ACKNOWLEDGMENTS

The submitted paper has been authored by a contractor of the U.S. Government under contract no. W-31-109-ENG-38. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. Work supported by the U.S. Department of Energy, Office of Environmental Management, under contract W-31-109-ENG-38.

REFERENCES

Autodesk, Inc. (1993) *AutoCAD Release 12 - Advanced Tools*, pp. 267-315.

ConSolve, Inc. (1995) *SiteLink File Transfers*, Wayland, MA.

National Institute for Standards and Technology (1992) *FIPS Publication 173: Spatial Data Standard*, U.S. Department of Commerce.

ESRI, Inc. (1995a) *SDTS: Supporting the Spatial Data Transfer Standard in ARC/INFO*, Environmental Systems Research Institute, Inc. (Available from URL http://www.esri.com/resources/papers/papers.html.)

ESRI, Inc. (1995b) *ArcView: Shapefile Technical Description*, Environmental Systems Research Institute, Inc. (Available from URL http://www.esri.com/resources/papers/papers.html.)

NIST: See National Institute for Standards and Technology

AUTHOR INFORMATION

James A. Kuiper: Biogeographer / GIS Analyst
Environmental Assessment Division - 900/D11
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4832
Telephone: (708) 252-6206 FAX: (708) 252-3659
Internet: kuiper@spatial.ead.anl.gov

Co-Authors:
Andy Ayers and Robert Johnson
Argonne National Laboratory / Environmental Assessment Division

Marilyn Tolbert-Smith
Office of Environmental Management
U.S. Department of Energy

A black and white Postscript version of this paper (as submitted to the conference on 12/15/95) is available as an uncompressed file (5.7 Mb), a Sun compressed file (1.1 Mb) or a GNU zipped file (1 Mb).

DISCLAIMERS

---

http://www.ncgia.ucsb.edu/conf/sf_papers/kuiper_jim/sitelink.html

February 1, 1996

Trish Duce, Ellen Voth and Patricia Andrews

Visualization of Historical Wildfire Data: Application of a DX-Oracle Interface

A link between the Oracle database management system and IBM's Data Explorer (DX) scientific visualization software was developed. DX tools provide the capability to create more sophisticated visualizations than may be possible with a Geographic Information System (GIS). The technology is demonstrated on visualization of historical wildfire data from an Oracle database overlayed on data layers from a GIS.

Introduction

Historical wildfire data are used for a variety of applications including post-season analysis of fire activity and evaluation of fire danger rating indexes. Fire location according to administrative unit (e.g. Forest or Wilderness area) is sufficient resolution for some applications. There is location information (latitude and longitude) in the Forest Service wildfire database, however, that has not been used to full advantage. Forest Service Regions produce annual fire reports summarizing the fire activity for each year (e.g. USDA Forest Service 1994). The report includes tables such as number of fires and acres burned for lightning and person-caused fires for each Forest. A visualization that displays both time and space components of fire activity would supplement summary tables.

Fire occurrence and size are related to fire danger rating indexes to evaluate the effectiveness of the National Fire Danger Rating System (Deeming and others 1977). The relationship between fire activity and fire danger rating is used to evaluate severity of the fire season (Bradshaw and Andrews, in press) and also to better use fire danger rating in decisionmaking (Andrews and Bradshaw 1995). Because the current fire danger rating system is based on weather observations taken at a fixed station and applied to a large area, fire location according to administrative unit is appropriate. The next generation system, the Wildland Fire Assessment System (WFAS), will be based on spatial data for fuel, fuel moisture, and weather (Andrews and others, in press). Using fire location according to latitude and longitude is critical for evaluation and use of WFAS.

In this paper we describe demonstration of software that was developed to extract data from an Oracle data base and display it with Data Explorer (DX) scientific visualization software. Wildfire data are displayed on elevation, land cover classification, and a biweekly live fuel greenness layer, providing a visualization of the fire activity as it varies in both space and time.

Wildland Fire Databases

USDA Forest Service historical wildfire data (1970-present) is stored in the National Interagency Fire Management Integrated Database (NIFMID), an Oracle database at the USDA National Computer Center in Kansas City (USDA Forest Service 1993). Of the information included for each fire, only the following is used in this demonstration: fire identification number, ownership (region and forest), location (latitude and longitude), discovery date (month/day/year), statistical cause (lightning or human), and final fire size. The fire location information is recorded in latitude/longitude from 1986 to present.

Several data layers are used in conjunction with the fire data for display: elevation, land cover classification, and a biweekly greenness index. The land cover characterization map is available for the conterminous 48 States of the U.S. (Loveland and others 1991). It is derived from AVHRR satellite data and is raster data at 1 km resolution. Land cover classification is being interpreted as fuel type for wildfire (Burgan and Hardy 1994). The greenness maps, which indicate state of live vegetation, are also derived from AVHRR satellite data (Burgan and Hartford 1993). The 1 km resolution maps are updated and made available to the field on a weekly basis. Historical images are available from 1989 to present.

Other ancillary data from a GIS data base that can be used for reference in visualization of fire location include administrative boundaries (state, county, land ownership), lakes and rivers, roads, and cities. In this demonstration, we use only state and county boundaries.

Oracle Database Management System

Oracle is a modern relational database management system. It acts as an interface between the physical storage and the logical presentation of data. It provides a set of sophisticated tools for handling information. If you want to access and manipulate Oracle data, you need SQL - Structured Query Language. SQL has become the database language of choice because it is flexible, powerful, and easy to learn.

An Oracle Precompiler is a programming tool that allows you to embed SQL statements in a high-level source program. The precompilier accepts the source program as input, translates the embedded SQL statements into standard Oracle runtime library calls, and generates a modified source program that you can compile, link, and execute in the usual way.

When designing the DX-Oracle interface it was obvious there was a need to access the Oracle data. Pro*C was the Oracle Precompiler choosen for this project. Therefore the language C was used with embedded SQL commands to construct the DX-Oracle interface.

Data Explorer

Data Explorer, or DX, is a software package developed by IBM for scientific data visualization and analysis. It provides an extensive library of "modules" -- already written blocks of code which can be tied together into visual programs through the graphical user interface. The user 'writes' a visual program for rendering an image, based on the data, by choosing and connecting modules through the click and drag graphical program editor.

DX is similar to a GIS, in that it can map data to geographic locations, but in addition, it offers new capabilities for investigating large quantities of complex data. It can be used to explore geographic data sets but has also been used to visualize data from disciplines as varied as physics, geology, medicine, and meteorology. The aspects of DX which allow it this flexibility are described below. DX is:

Multidimensional - 3D objects can be created and manipulated just as easily as 2D plots. Other parameters or variables can be mapped onto objects as another dimension through the use of color, 'glyphs', or 'isosurfaces', thus allowing visual discovery of juxtapositions, continuities or irregularities.

Interactive - User interactivity allows the scientist to explore the data more completely. For example, created visual objects can be rotated to allow different points of view or the user can interactively modify the imagery by changing input values.

Temporal - DX allows time sequencing or animation so temporal processes can be viewed directly.

Modular - DX allows various levels of sophistication. Generic software building blocks allow a user to create a visual program or 'network' without ever writing a line of code. Although programming skills can be utilized in converting data files to DX format, and in creating new modules for use within the system they are not necessary to create visualizations of complex data.

A simple visual program is pictured below in Figure 1.

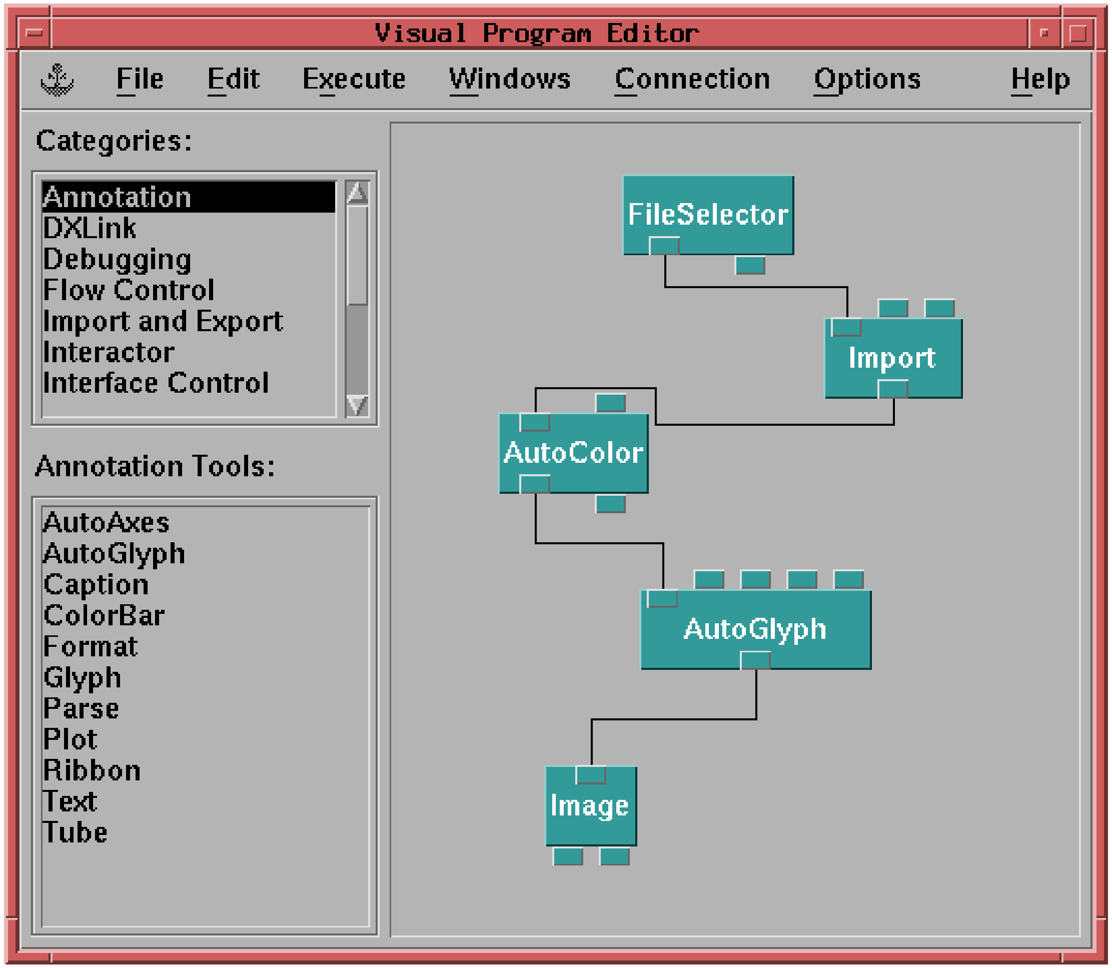
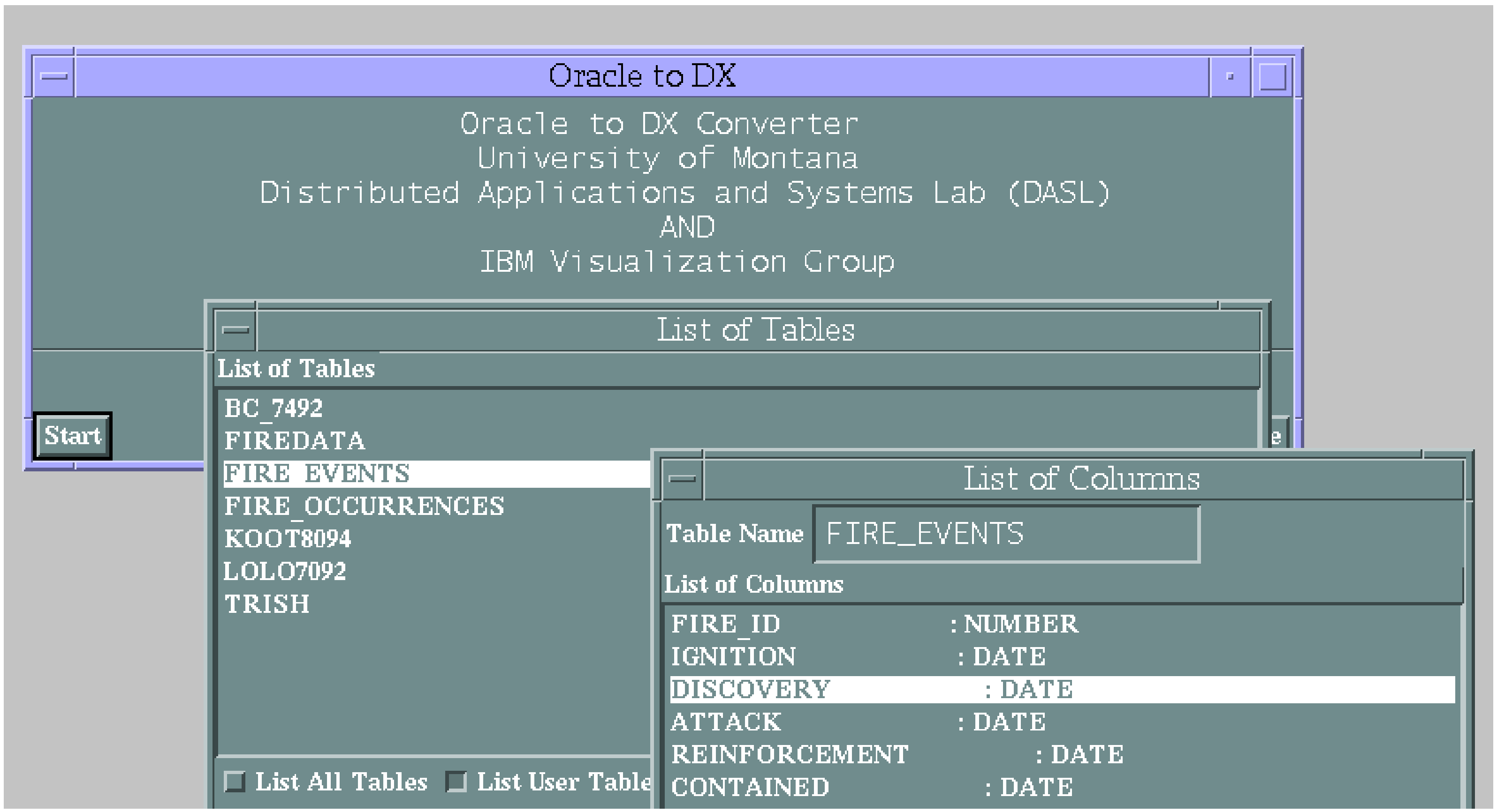


Figure 1. Example DX Visual Program

This network imports data from a file, colors and 'glyphs' an object based on the data values and then renders the object in an image window. This same network can be used to depict different data sets, e.g., carbon monoxide levels measured at different locations or the position and size of fires burning on a landscape. A sequencer module could be added to animate the image and view how the values change over time.

DX-Oracle Interface

An interface between DX and the Oracle database management system has been developed and implemented. The graphical user interface allows the user to connect to oracle and retrieve data from the database. The user has the option of scrolling through all of the tables of the database or just those tables the user owns. When the user double-clicks on a table, a box with the specified table's columns appears. The user can scroll through this list and select any columns he or she wants to retrieve. This can be done with multiple tables depending on what information the user is trying to retrieve. The user has the option of retrieving a subset of the columns he or she has selected by specifying a condition. Once the tables, columns and conditions are specified the user clicks on the "retrieve" button and a window pops up that asks the user to specify a file to save the information in. Once this has been done a window with the select statement created by the user appears. The user can press "OK" and retrieve the information or can press "cancel" if there was an error.



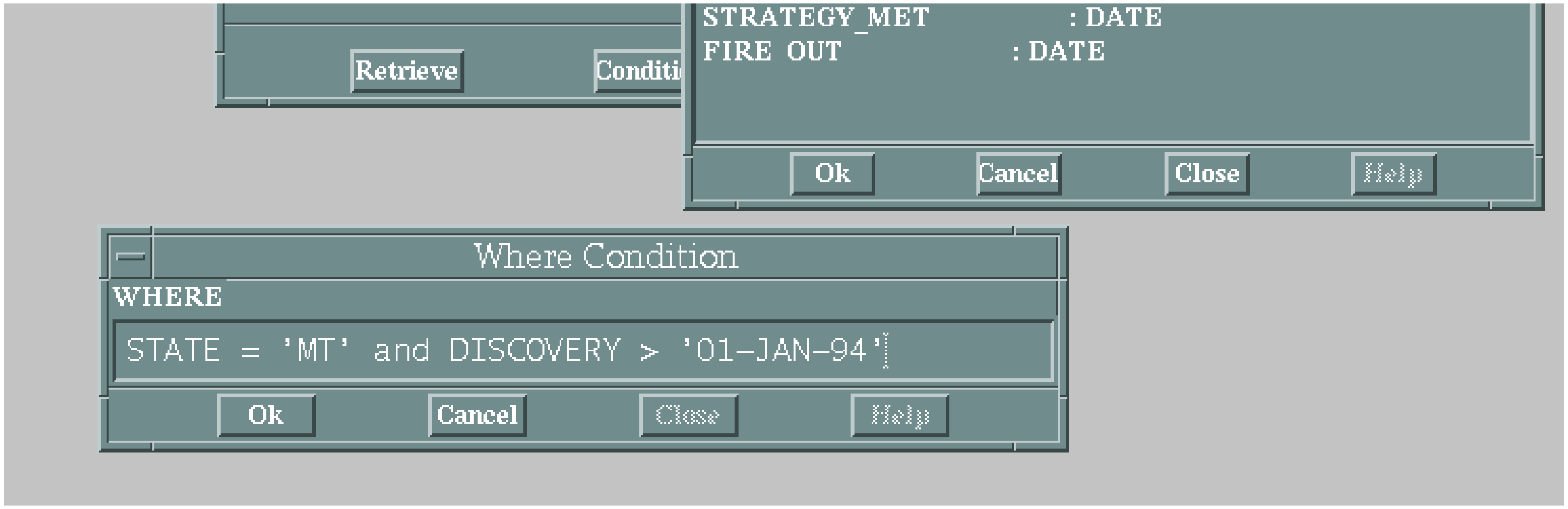


Figure 2. Oracle to DX Converter

The output of the program is a DX file that has an object for each column the user specifies. A dummy positions component is created so DX will load the data without any problems. It is up to the user to create a positions component in DX. If an Oracle column is of type "date" the program will automatically generate two objects. One with the actual date values and one with the same date values in the Julian format.

Changing Projections

One of the hardest parts of visualizing the historical fire data generated by the DX-Oracle interface was referencing it with other data sets. The fire's location is recorded in geographic longitude and latitude coordinates. All of the data sets we wanted to reference the fire data to were in cartesian coordinates. Cartesian coordinates come in many different projections. Digital Line Graphs (DLGs) and Digital Elevation Maps (DEMs) were the two types of data sets used to create visualizations with the fire data. Gerald I. Evenden had developed Cartographic Projection Procedures for the UNIX environment. These procedures will convert geographic longitude and latitude coordinates into many cartesion projections and many of the inverse conversions can also be performed. A Data Explorer module that uses the cartographic projection utilities was created by Dave Thompson at the University of Montana. This module was used to convert the fire data into cartesian coordinates.

Fire History Visualization Demonstration

The states of Montana and Arizona were chosen as demonstration sites. The first visualization uses the discovery dates and fire out dates to show the number and location of fires burning on any particular day in Montana for the year 1988. The background is colored and scaled according to elevation (Figure 3). The second Montana visualization is a comparison of the 1993 and 1994 fire seasons using simultaneous display of two maps with glyphs indicating fire occurrence (based on fire discovery date) and size throughout the season. The background is colored according to elevation. A plot of number of fires is updated as the display of the fire season progresses (Figure 4). The two seasons are very different, 1994 being one of the most active on record. The visualization is an effective display of that difference.

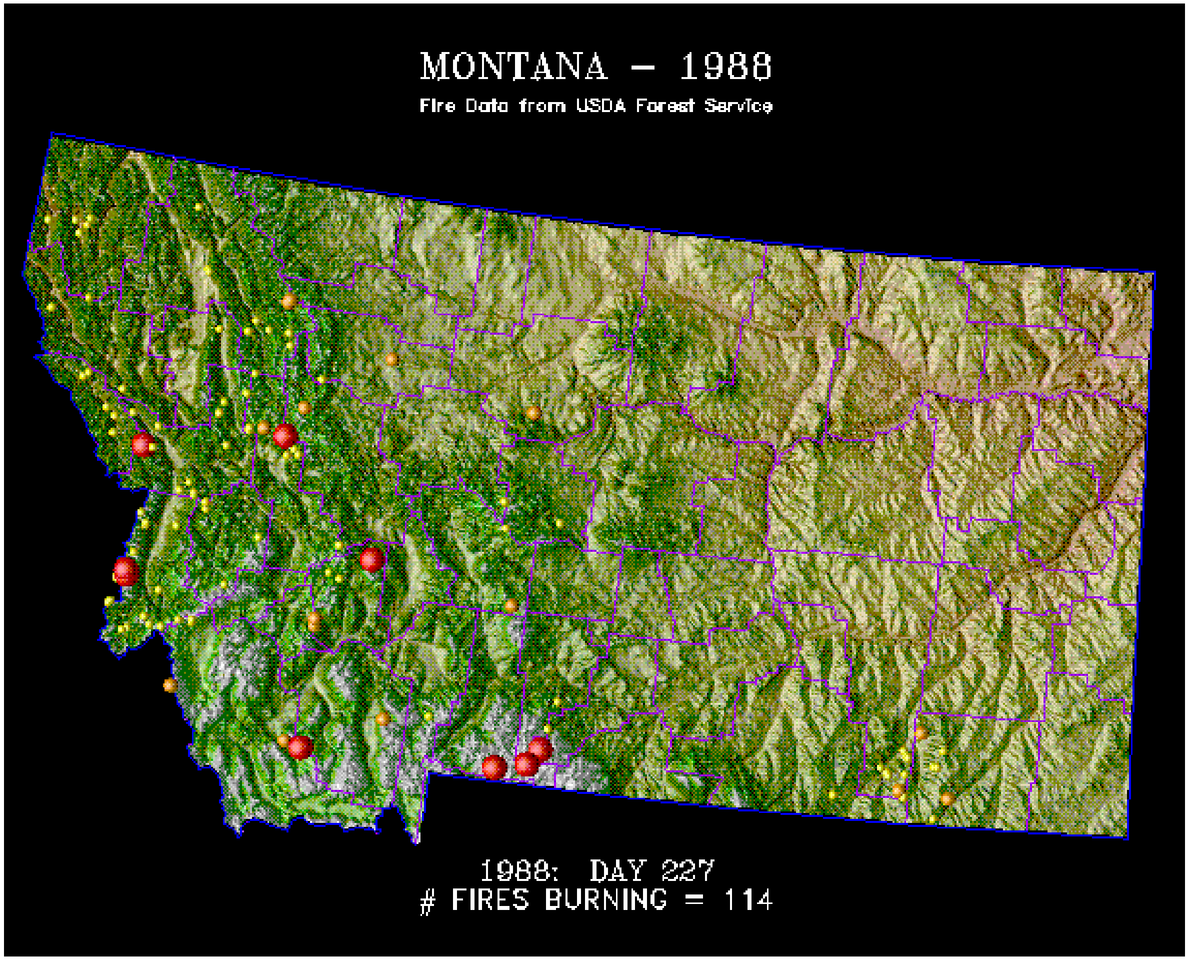
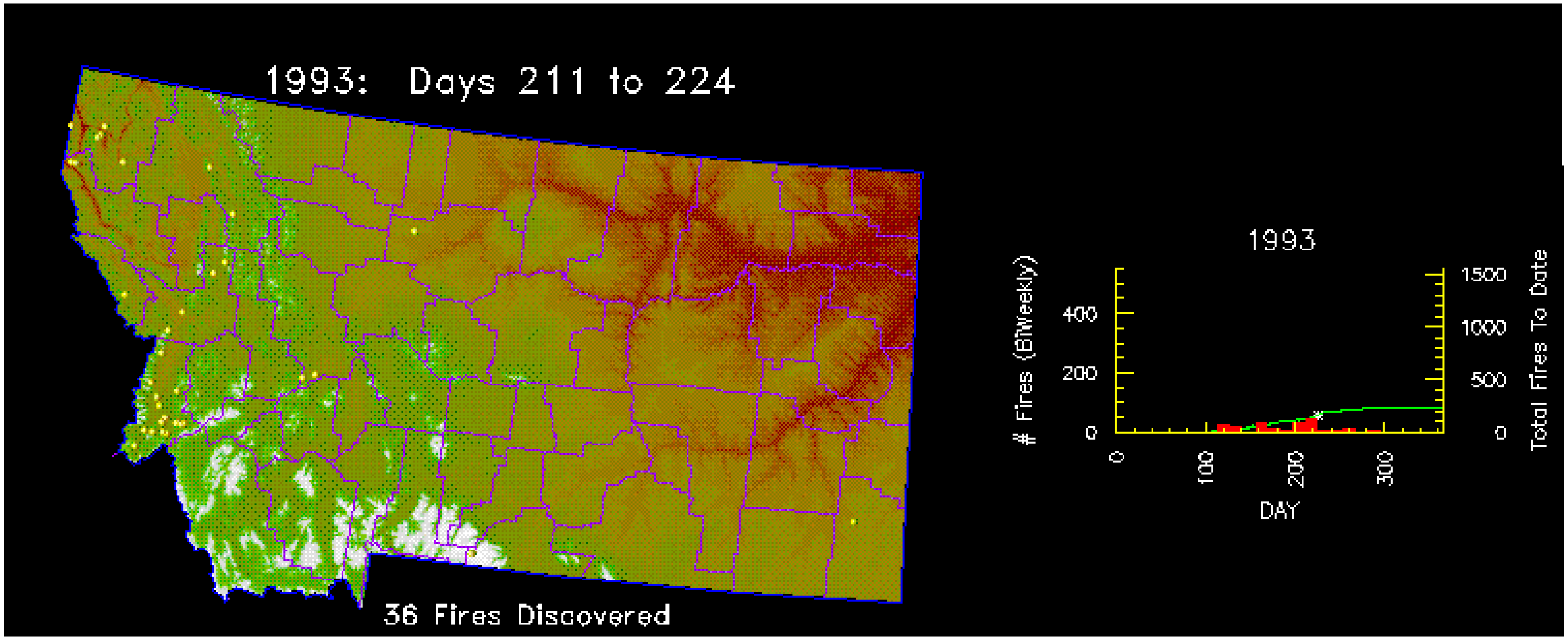


Figure 3. Wildfires Burning in Montana on August 14, 1988



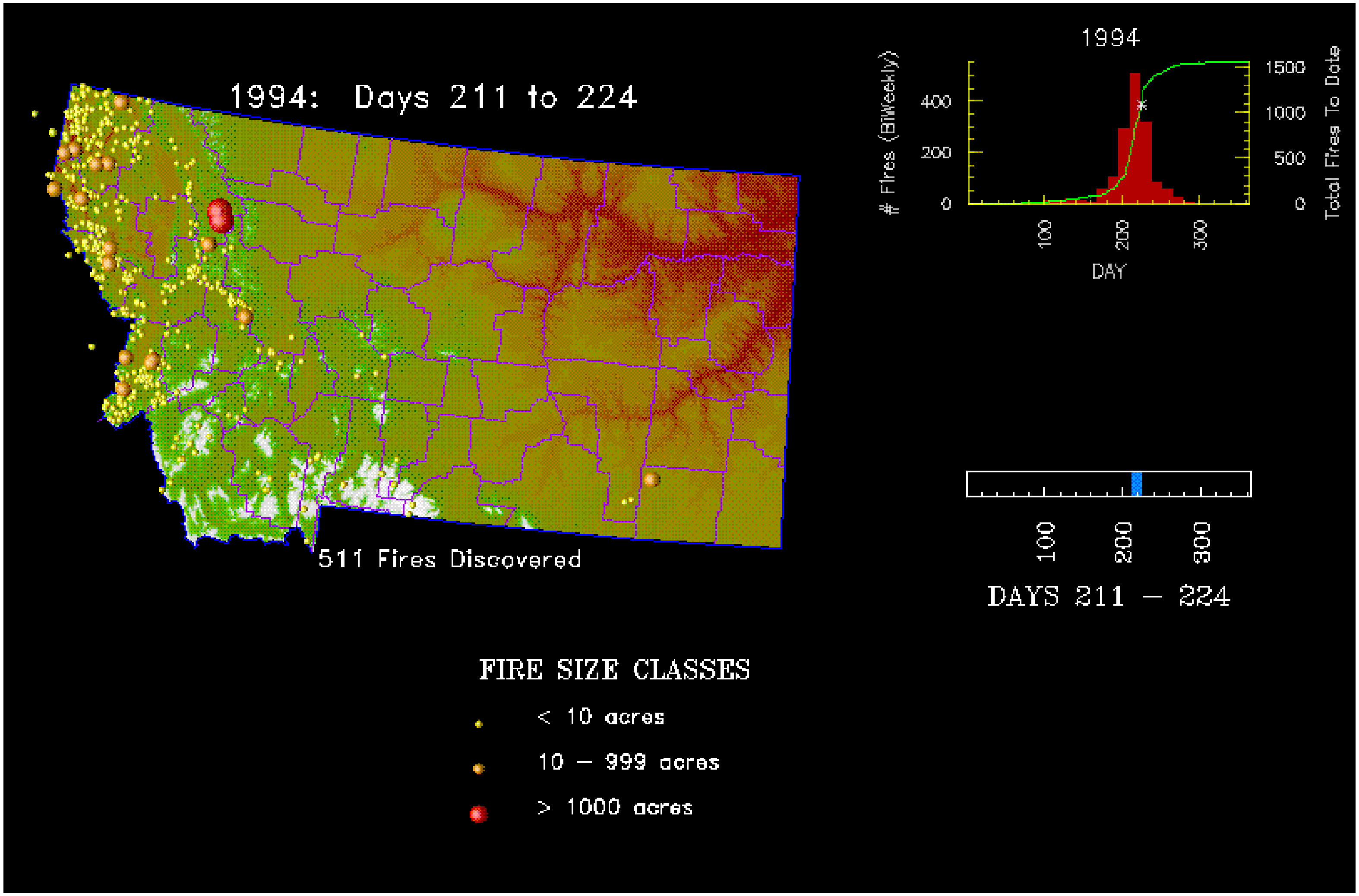
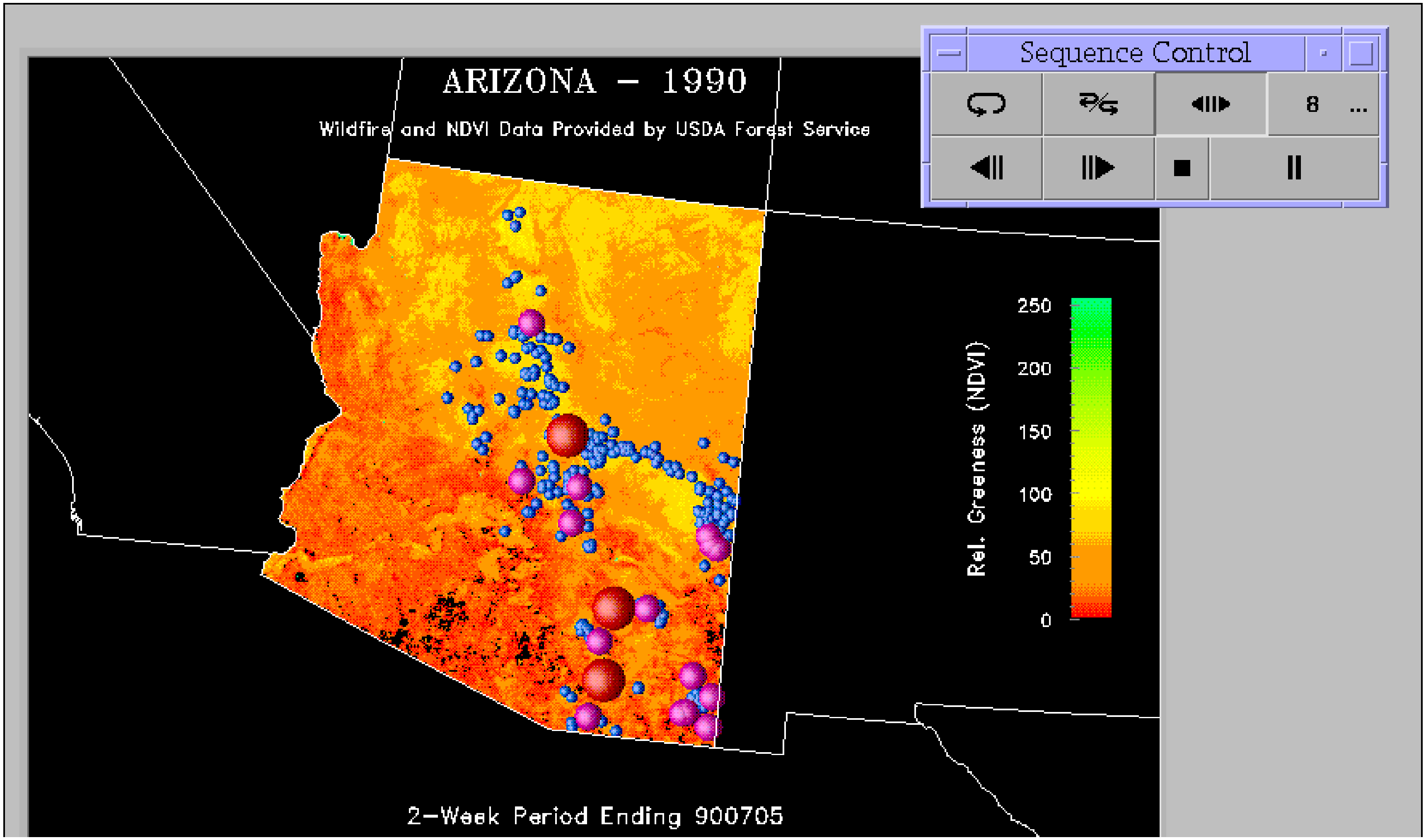


Figure 4. Wildfires Discovered in Montana from July 29 thru August 11 for 1993 and 1994

The background for the Arizona fire data is a greenness map derived from satellite data. It is an indication of fire potential and changes biweekly. Fire location and size is displayed in two week blocks corresponding with the change in greenness (Figure 5). The animation of greenness and fire activity is animated for March through October of the year selected (1990-1993). Developers of the next generation fire danger rating system will be able to use this as the first step in assessing the relationship between fire activity and spatial indicators of fire potential.

This demonstration showed the success of a link between Oracle and DX and of the benefit uf using DX to visualize historical fire data in conjunction with GIS data layers.



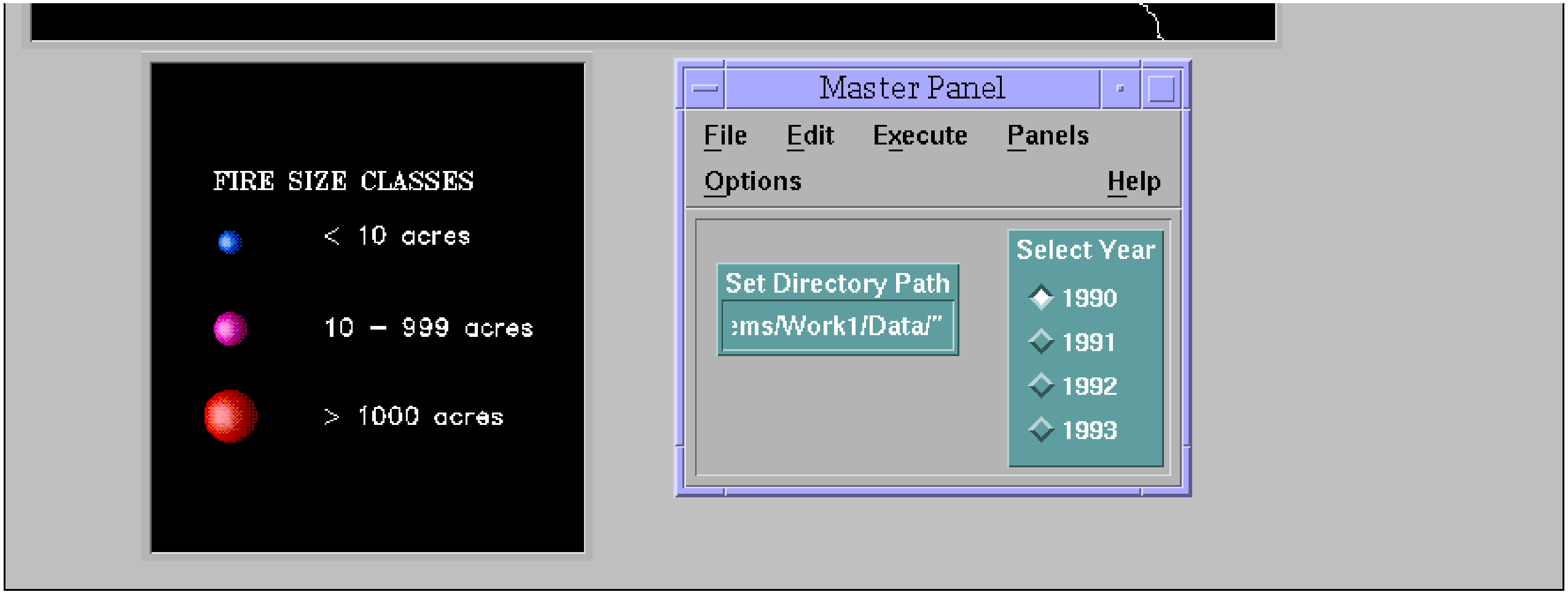


Figure 5. Arizona Wildfire Occurrence Plotted on a Greeness Map for the Two Week Period Ending July 5, 1990

Acknowledgments

Software development and visualizations were done by Trish Duce and Ellen Voth under an internship program with IBM in Boulder Colorado.

Vijayant Palaiya from the University of Montana contributed to implementation of the DX-Oracle user interface.

References

Andrews, Patricia L., and Bradshaw, Larry S. (1995) Fire Danger Rating and the Go/No-Go Decision for Prescribed Natural Fire. Proceedings:symposium on fire in wilderness and park management; Missoula, MT, March 30-April 1, 1993. General Technical Report

INT-GTR-320. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. pp. 149-151.

Andrews, Patricia L., Bradshaw, Larry S., Burgan, Robert E., Chase, Carolyn H., and Hartford, Roberta A. (in press) WFAS: Wildland Fire Assessment System--Status 1995, Presented at 1995 Interior West Fire Council Meeting, St. George, Utah. Nov. 1-3, 1995.

Bradshaw, Larry S., and Andrews, Patricia L. (in press) Fire Potential Assessment During a Period of High Fire Activity in the Northern Rockies: August 1994. A paper presented at Interior West Fire Council, Coeur d'Alene, ID, November, 1994.

Burgan, Robert E., and Hardy, Colin C. (1994) Ground Truthing a National AVHRR Based Vegetation/Fuels Map. A paper presented at the 12th Conference on Fire and Forest Meteorology, Jekyll Island, GA. October 26-28, 1993.

Burgan, Robert E., and Hartford, Roberta A. (1993) Monitoring Vegetation Greenness With Satellite Data. Gen. Tech. Rep. INT-297. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 13 P.

Deeming, John E., Burgan, Robert E., Cohen, Jack D. (1977) The National Fire Danger Rating System-1978. General Technical Report INT-39. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 63 p.

Evenden, Gerald I. (1991) Cartographic Projection Procedures for the UNIX Environment--A User's Manual..

Loveland, Thomas R., Merchant, James W., Ohlen, Donald O., and Brown, Jesslyn F. (1991) Development of a Land-Cover Characteristics Database for the Conterminous U.S. Photogrammetric Engineering and Remote Sensing, 57:11, 1453-1463.

Perry, James T., Lateer, Joseph G. (1992) Understanding Oracle. BPB Publications.

Programmer's Guide to the Oracle Pro*C Precompiler. Release 2. Chapter 1.

USDA Forest Service. (draft, 1993) NIFMID: National Interagency Fire Management Integrated Database- Technical Reference Manual. U.S. Department of Agriculture, Forest Service. Fire and Aviation Management, Washington, D.C.

USDA Forest Service. (1994) 1994 Annual Fire Report U.S. Department of Agriculture, Forest Service, Northern Region, Missoula, MT, 33 p.

Trish Duce
University of Montana
Missoula, MT 59812
Telephone: (406) 243-5779

email: trish@cs.umt.edu

Ellen Voth
University of Montana
Missoula, MT 59812
Telephone: (406) 543-4622
email: evoth@selway.umt.edu

Patricia Andrews
Intermountain Research Station
Forest Service, U.S. Department of Agriculture
Missoula, MT 59807
Telephone: (406) 329-4827
FAX: (406) 329-4825

Individual-Based Models in Ecology: An Overview

D. L. DeAngelis, National Biological Service, University of Miami, Coral Gables, FL 33124
D. M. Fleming, National Biological Service, Everglades National Park, Homestead, FL 33034
L. J. Gross, University of Tennessee, Knoxville, TN 37996
W. F. Wolff, Forschungszentrum Juelich, Juelich, Germany

Abstract: Individual-based population models simulate the behavior of each member of a biological population as an individual. These models differ from traditional state variable models in which population size is described as an aggregated variable. The advantages of individual-based models include the following:

(1) A variety of types of differences among individuals in the population can be accommodated
(2) complex decision making by individual can be simulated
(3) local interactions in space and the effects of stochastic temporal and spatial variability are easily handled.

In addition to these advantages, individual-based models can be combined with geographic information systems to address applied problems. Currently, this approach is being used to model the biotic communities of the Everglades landscape. Several individual-based model of higher trophic level species are being developed and integrated with models of abiotic variables such as the seasonally varying water levels in the Everglades, and with process models of lower trophic levels across the heterogeneous Everglades landscape. The methodology for linking these models through GIS is described. Implications for environmental assessment are discussed.

Contact: Donald L. DeAngelis
Ecologist with the National Biological Service
Department of Biology
University of Miami
PO Box 249118
Coral Gables, FL 33124-0421
phone: 305-284-1690
fax: 305-284-3039
ddeangelis@umiami.ir.miami.edu

Timothy A. Kohler, Carla R. Van West, Eric P. Carr, and Christopher G. Langton

Agent-Based Modeling of Prehistoric Settlement Systems in the Northern American Southwest

Archaeologists have long been interested in why ancient peoples located their settlements where they did and why these settlements were abandoned. These questions have always been approached by comparing known site distributions with resource distributions of various types, possibly (but not often) taking into account how resource distributions might have been different in prehistory. It has never been possible, with these techniques, to also model the effect of humans on the landscapes they occupy. Recently, these inductive settlement pattern studies have been greatly aided by GIS and a growing battery of statistical techniques. Here we report the first stages of a long-term project to begin understanding settlement processes by departing from household-level decision-making rather than from the archaeological record. We simulate the placement of residences and of population growth and decline, as they respond to changing maize production opportunities and local human impact on the environment, in southwestern Colorado between A.D. 900 and 1300. The effort combines GIS data planes and agent-based modeling in a way that seems promising for many observational sciences.

See also:
http://www.santafe.edu/~carr/model/village.html
http://www.santafe.edu/projects/swarm

INTRODUCTION

Despite their many differences, sciences that share a strong observational component also share the serious problem of having to infer process from pattern. Some traditionally observational sciences, including ecology, have been able to introduce experimentation as one way to help overcome this difficulty. Our ability to do this in archaeology, however, is severely limited. There are even problems with applying the weaker procedure of building analogies, for many of the societies in which we are interested may have no very applicable living or historically documented analogs. Finally, we have to cope with the fact that many things that we would like to know about prehistoric societies--for example, their kinship systems, religious practices, and political and economic traditions--are not directly preserved in the archaeological record.

In the spirit of trying to start with what can be observed in the archaeological record and working towards what can not, many archaeologists have invested a great deal of effort in attempting to understand prehistoric settlement patterns. Except where the tempo of human movement across the landscape was rapid and non-repetitive in pattern, archaeologists have a good chance of finding and dating human settlements. Archaeologists have been particularly successful in building accurate and complete maps of human settlement for the late prehistoric period in the U.S. Southwest, where relatively sedentary farmers left frequently substantial remains that can often be dated with relative precision. The resultant settlement maps can be, and have been, compared with maps of known resource distributions to describe, through statistical inference, the relative strengths of various resource distributions in affecting human settlement (Kohler and Parker 1986:402-431 provide an introduction to such work). Similar work around the world has been greatly aided by GIS in the last decade.

The comparative completeness and accuracy of these maps, however, has not magically dissipated various problems of inferring the processes that produced these settlement patterns �if anything, it has made the difficulty of carrying out this inferential procedure in a satisfactory way even more embarrassing. It gets harder and harder to blame the foibles of the archaeological record for our inability to agree on the processes that produced it as that record becomes increasingly complete! Key problems for the traditional inferential procedure have included:

- the difficulty of making the static maps of resource distributions, as displayed for example by a GIS, responsive to known changes in climatic conditions during the period of occupation;
- the difficulty of making the maps of resource distributions responsive to their use and alteration by people; and
- the difficulty of effectively incorporating aspects of the contemporary social environment in the mix of variables considered to affect human settlement locations.

In short, traditional inferential approaches fail to capture the dynamic and coevolutionary nature of human settlement decisions as they respond to shifting resources and the presence and actions of other people. As a result, to take as example the Anasazi region where the project we are reporting is located, there remain large disagreements among archaeologists concerning the relative importance of warfare, political competition, climate change, and human impact on the environment in effecting the settlement patterns we observe. If only we had a laboratory to study the outcomes of various processes as they might play themselves out through hundreds of years on realistic landscapes! This project reports our current progress towards building just such a virtual laboratory, and outlines our eventual goals.

Agent-Based Modeling on Dynamic Landscapes

Although archaeologists since the 1960s have made some use of simulation, most such work has been at the systems level. Well known examples include the attempt to apply systems-dynamic simulation in the spirit of Forrester (1968) to the problem of the Classic Maya collapse (Hosler et al 1977). By their very nature, such simulations failed to capture the importance of space in conditioning type and frequency of social interaction, and, by default, considered societies to be either internally homogeneous or characterizable by reference to a very few discrete internal components (such as elites and commoners).

A few years ago a handful of researchers in anthropology began the difficult task of modeling societies on the computer from the bottom up, beginning with the individual or the household;

a notable example is the work by Jim Doran et al. (e.g., 1994) on Upper Paleolithic societies (see also Biskowski [1992] and Renfrew [1977] for more general programmatic statements). British researchers often discuss such models as examples of "distributed artificial intelligence," emphasizing that they represent cognitive processes at the level of the individual. Similar approaches are called individual-based models in the ecological literature and agent-based models by many U.S. social researchers. Regardless of what they are called, these models share an emphasis on modeling behavior at the lowest practical level, and an interest in studying the emergence of spatial arrangements and interactions among agents, and the nature and evolution of strategies for agent interaction with the environment and with other agents. Whereas systems-level models often focused on finding parameter ranges that would permit system homeostasis, agent-based models tend to focus on the problem of morphogenesis (change in structure of the macro-system) as a result of interactions at the micro level.

The Present Project

The immediate motivation for the present project was a desire to understand why, during certain times in prehistory, most Puebloans lived in relatively compact villages, while at other times, they lived in dispersed hamlets (Cordell et al. 1994). Our chosen approach has roots extending at least 15 years back into the early 1980s, when a dissertation from the University of Arizona by Barney Burns (1983) showed that it was possible to retrodict potential prehistoric maize y ields in a portion of southwest Colorado by combining prehistoric tree-ring records with historic crop-production records of local farmers. A few years later, Kohler et al (1986; see also Orcutt et al. 1990) simulated agricultural catchment size and shape in a northern portion of the present study area, to arrive at the suggestion that avoiding violent confrontation over access to superior agricultural land was a major force in forming the villages that appeared in this area in the late A.D. 700s and again in the mid 800s. Shortly after that, Carla Van West, in a 1990 dissertation (published 1994) used a different and larger set of tree-ring data to produce spatialized Palmer Drought Severity Indexes in 1,070 GIS data planes, one for each year from A.D. 900-1970, for the portion of Southwestern Colorado, at a spatial resolution of 4 ha. Construction of these landscapes is described briefly below. Van West was the first to use a series of local weather stations and specific soil types to reconstruct PDSI in a way that made these measures respond to very local conditions. Finally, Kohler and Van West (1996) examined these production landscapes against the known record of aggregation in this area and suggested that microeconomic processes�at the level of the household�could successfully explain whether settlement was dispersed or aggregated at any time. Specifically, we suggested that villages formed during periods when it was in the best interests of households to share food with other households, and dissolved when it was in the best interests of households to hoard their production. Our arguments were based on comparing the relative payoffs to households of sharing vs. hoarding under various production regimes, using sigmoid-shaped utility curves. To test this model more rigorously, however, we needed to simulate household placement, maize production, consumption, and exchange with other households in considerable detail.

Constructing the Paleoproduction Landscapes

The 400 yearly maps of potential agricultural potential used in this simulation were produced as follows.

- First, a study area was defined. The land area selected had to have sufficiently diverse landforms, variable elevational settings, and well-documented archaeological sites dating from the appropriate time periods. Further, it had to exist as 7.5-minute DEMs. Eventually, Van West selected an area equivalent to 12 7.5-minute topographic maps to represent the maximum study area.
- Second, a base map for the study was produced by mosaicking the 12 DEMs into a single image.
- Third, the spatial resolution of the analysis was established. The image was partitioned into rows and columns, where each cell was 4 ha (200 m by 200 m) in area. Subsequent to trimming, the final image was organized as 200 rows and 227 columns (DEMWIND).
- Fourth, Van West recorded selected attributes for each of the 45,400 cells in the image. Soil data in the form of soil series information, soil depth, available water capacity, and natural plant productivity, as well as agricultural yield information, were recorded for each cell. The soil series information was used to create a second data layer or map which depicted the distribution of 98 distinct soil types (SOILWIND). A total of 36,759 cells representing 81% of the study area (1,470 km2) had complete soils information and was used in all subsequent steps.
- Fifth, derivative maps that reduced the elevational and soils data into more interpretable forms were produced. A third map (11AWC) was derived that depicted the 98 soil types classified into one of 11 soil moisture classes, which could be used to calculate Palmer Drought Severity Index (PDSI) rankings. A fourth map that classified the 1,512 different elevational values into one of five elevational bands also was created (5ELEV). Each elevational band was associated with the instrumented records of an appropriate and proximate weather station (Bluff, Utah; Cortez, Ignacio, Mesa Verde, and Fort Lewis, Colorado).
- Sixth, the climatic and the soils data were used to calculate PDSI values. These took the form of monthly precipitation and temperature data from the five, elevationally diverse weather stations, as well as data on the available water capacities and soil depth for the 11 contrasting soil-moisture classes. PDSIs are temporally sensitive indicators of soil moisture; they are commonly used to model the success of dry farming agriculture. Negative values indicate dry conditions, and positive values indicate moist conditions. As with tree-ring width data, PDSIs integrate the effects of precipitation and temperature on available stores of soil moisture and incorporate the balance from previous months into the estimate for the current month. In this study, PDSIs calculated for the month of June during the historic instrumented period were correlated with tree-ring width data for the same set of years. This produced a calibration, or transfer function, that was applied to the full length of the tree-ring series, A.D. 901 to 1970. In this way, 55 long-term reconstructions of PDSI were produced, one for each combination of the five elevational strata with the 11 soil moisture groups.
- Seventh, Van West created a fifth map illustrating the precise spatial distribution of each of the 55 long-term reconstructions (PDSIMAP). It combined data from 11AWC and 5ELEV.
- Eighth, the 55 long-term reconstructions were used to assign annually specific PDSI values to each cell in the PDSIMAP image. Consequently, 1070 data planes were created and represented a continuous record of potential soil moisture conditions for the

A.D. 901 through A.D. 1970 period. Each summarizes spatially variable soil moisture conditions as they existed across the study area on July 1, just before the advent of typical summer rains.

- Ninth, all the PDSI values compiled for a single year were reexpressed as potential crop yields (first as beans and later as maize) and summed. This was accomplished through regression analysis and estimation. The relationship between natural plant productivity data and historical crop yield data for 44 of the 98 soil types in the study area were used to estimate potential yield on soils without crop yield data. The end products of these GIS-coordinated steps were the creation of 400 annual maps (PRODyyyy from A.D. 901 to 1300) depicting the distribution of climatically conditioned yield values for potential maize production and a tabular summary of yearly yield estimates for the four century-period. In general, the EPPL7 GIS system was used to produce, coordinate, and display these data, and a variety of statistical packages (SPSS, SAS, SYSTAT) were used to calculate numerical values for subsequent analyses.

Present Status of the Modeling Effort

The village simulation based on these landscapes is being developed using the Swarm agent simulation libraries now under design at the Santa Fe Institute. The primary focus of the current model is to create plausible agent behavior on the paleoproductivity landscapes, where agents represent households (the minimum observational unit in the local archaeological record). This involves endowing our agents with realistic heuristics that specify how households might have reacted to varying planting environments from year to year. Currently, these decisions are static.

The basic structure of the model is a rectangular lattice representing the area described above. Each "cell" in this model consists of a 4-ha square which keeps track of relevant information from the paleoproduction database. Upon this landscape exist numerous households.

Each household makes planting decisions based upon past expectations of harvests and fertility of land gathered from local cells around their central location. Dependent upon the past success of the agent's strategy in the local environment, the household might opt to search a wider area and possibly relocate if the internal storage of maize is dangerously low. The household also has probabilistic natality and mortality rules specifying when members are born and die. New household formation or "marriage" provides a dynamic element to planting and location considerations since neighbors will influence how much available land there is for planting in any local environment. Efforts are now focused on providing a plausible model of how decision-making concerning planting and residence location on this landscape. Onto this model will be added exchange relationships, more complex learning rules, additional household activities, etc. over the course of the coming year.

The primary area of research will be the effect of exchange relationships upon the formation of larger social groups. Since agricultural yields varied greatly from year to year, farmers needed to adapt mechanisms to reduce their uncertainty of future yields. One such mechanism thought to be important is reciprocity between households. After a reasonable model of agent planting is constructed, we will endow agents with balanced reciprocity behaviors and adaptive encodings of exchange, placing the households into a social and an economic network or other (related and unrelated) households. This network will be flexible enough to evolve according to agent interactions and changes in the world environment.

Longer-term Goals

Among our longer-term goals is the desire to analyze how different agent behaviors influence the model. Providing heuristic behaviors that represent plausible social norms and comparing results with rational expectation type optimizing behavior (and comparing the outcomes of each with the known archaeological record) will help us understand the extent to which such behaviors were in fact optimized, and with respect to what. The larger Swarm project at the Santa Fe Institute, of which this effort is a small part, is also working to provide mechanisms to allow agents to run simulations of the world, given their internal beliefs about the world, in order to make decisions for action. This ability would allow households, in effect, to "imagine" the probable outcomes of various decisions and select behaviors based on this anticipation, capturing at least part of what makes human decision-making so flexible and difficult to analyze. Making optimizing decisions endogenous to the model through use of genetic algorithms or classifier-type learning (e.g., Reynolds 1986) could conceivably produce results comparable with the historical record.

Conclusions

It is not the goal of this project, however, to generate settlement patterns identical to those found by archeologists in the Four Corners area of the United States, although the extent to which our simulated settlement patterns approach those in fact found should help us understand the processes that generated those patterns. Just as important to us is the process of exploration and discovery. What agent-based behaviors can be used to explain emergent aggregate social structure? How can we model the behavior of emergent entities such as clans or villages where some decisions at the higher level displace or limit decisions made at the local level? We are entering into relatively uncharted waters where archaeology indeed becomes anthropology, but with the benefit of a data base accumulated over hundreds of years.

Acknowledgments

This work is partially supported by Grant No. MT-0424-5-NC-026 to Washington State University from the National Center for Preservation Training and Technology of the National Park Service, USA, and by the Santa Fe Institute, Santa Fe, New Mexico.

References

Biskowski, Martin (1992) Cultural Change, the Prehistoric Mind, and Archaeological Simulations. In *Archaeology and the Information Age: A Global Perspective*, edited by Paul Reilly and Sebastian Rahtz, pp. 212-229. Routledge, London.

Burns, B. T. (1983) Simulated Anasazi Storage Behavior using Crop Yields Reconstructed from Tree-Ring Records, A.D. 652-1968. 2 vols. Ph.D. Dissertation, University of Arizona.

University Microfilms, Ann Arbor.

Cordell, Linda S., David E. Doyel, and Keith W. Kintigh (1994) Processes of Aggregation in the Prehistoric Southwest. In *Themes in Southwest Prehistory,* edited by George J. Gumerman, pp. 109-133. School of American Research Press, Santa Fe, NM.

Doran, Jim, Mike Palmer, Nigel Gilbert, and Paul Mellars (1994) The EOS Project: Modelling Upper Paleolithic Social Change. In *Simulating Societies: the Computer Simulation of Social Phenomena,* edited by Nigel Gilbert and Jim Doran, pp. 195-221. UCL Press, London.

Forrester, J. (1968) *Principles of Systems.* Wright-Allen Press, Cambridge, MA.

Hosler, D., J. A. Sabloff, and D. Runge (1977) Simulation Model Development: A Case Study of the Classic Maya Collapse. In *Social Processes in Maya Prehistory,* edited by N. Hammond, pp. 552-590. Academic Press, London.

Kohler, Timothy A., and Sandra C. Parker (1986) Predictive Models for Archaeological Resource Location. In *Advances in Archaeological Method and Theory* 9, edited by M. B. Schiffer, pp. 397-452. Academic Press, Orlando.

Kohler, Timothy A., Janet D. Orcutt, Kenneth L. Petersen, and Eric Blinman (1986) Anasazi Spreadsheets: The Cost of Doing Agricultural Business in Prehistoric Dolores. In *Dolores Archaeological Program: Final Synthetic Report,* compiled by D. A. Breternitz, C. K. Robinson, and G. T. Gross, pp. 525-538. Bureau of Reclamation, Engineering and Research Center, Denver.

Kohler, Timothy A., and Carla R. Van West (1996) The Calculus of Self Interest in the Development of Cooperation: Sociopolitical Development and Risk among the Northern Anasazi. In *Evolving Complexity and Environment: Risk in the Prehistoric Southwest,* edited by Joseph A. and Bonnie B. Tainter, pp. 169-196. Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Vol. XXIV. Addison-Wesley, Reading, MA.

Orcutt, J. D., E. Blinman, and T. A. Kohler (1990) Explanations of Population Aggregation in the Mesa Verde Region prior to A.D. 900. In *Perspectives on Southwestern Prehistory,* edited by Charles Redman and Paul Minnis, pp. 196-212. Westview Press, Boulder.

Renfrew, A. C. (1987) Problems in Modelling Socio-Cultural Systems. *European Journal of Operational Research* 30:179-192.

Reynolds, R. G. (1986) An Adaptive Computer Model for the Evolution of Plant Collecting and Early Agriculture in the Eastern Valley of Oaxaca, Mexico. In *Guila Naquitz: Archaic foraging and early agriculture in Oaxaca, Mexico,* edited by Kent V. Flannery, pp. 439-500. Academic Press, Orlando.

Van West, Carla (1994) *Modeling Prehistoric Agricultural Productivity in Southwestern Colorado: A GIS Approach.* Reports of Investigations 67. Department of Anthropology, Washington State University, Pullman.

Timothy A. Kohler
Department of Anthropology
Washington State University
Pullman, WA 99164-4910 USA
ph: (509) 335-2698
e-mail: tako@wsu.edu

Carla R. Van West
Statistical Research, Inc.
P.O. Box 31865
Tucson, AZ 85751 USA
e-mail: sristats@aol.com

Eric P. Carr
Carleton College
Northfield, MI 55057 USA
e-mail: carr@santafe.edu

Christopher G. Langton
Santa Fe Institute
1399 Hyde Park Rd.
Santa Fe, NM 87501 USA
e-mail: cgl@santafe.edu

Modeling ecological patterns and processes using agent-based simulations and GIS

James H. Brown

Abstract:

I will talk about two collaborations in which colleagues and I have used simulation to investigate the structure and dynamics of complex, spatially explicit ecological systems. The first studies complement empirical studies that use ARCINFO GIS and the large data base of the North American Breeding Bird Survey (a standardized census that has been conducted annually since the mid-1960's and now includes about 2,000 locations across the US and Canada) to analyze spatial and temporal variation in the abundance of bird populations. The abundance of common species varies over several orders of magnitude and shows interesting patterns with respect to location within the geographic range. We are asking whether we can mimic and therefore begin to understand these patterns with computer simulation models of the niche combinations of environmental variables that limit abundance and distribution of each species. Using GIS and other data bases on weather, topography, soils, vegetation, etc., we can simulate the abundance and distribution of imaginary species with specified requirements. The other studies are computer-based experiments to understand the complex structures and dynamics produced by John Holland's artificial ecology ECHO. In this system agents ("individuals" of asexual "species") compete for resources, interact with other agents through "combat" (predation) and "trade" (mutualism), and reproduce with possible genetic change according to simple rules. The outcome of running ECHO for many generations is the evolution of an "ecological community" containing multiple species with divergent characteristics. By conducting controlled, replicated experiments in which we manipulate the environment (e.g., rate of resource supply) and/or the rules of agent behavior (eg., the kind, number, and strength of interactions), we can investigate the processes that affect "species diversity" and "community dynamics."

Contact: James Brown
Department of Biology
University of New Mexico
Albequerque, NM
dept: 505-277-9337
lab: 505-277-6437
email: jbrown@bootes.unm.edu
(Alternate contact: Sarah - 505-277-9337, mornings only)

Solving the Clear-Cut Scheduling Problem with Geographical Information Technology and Constraint Reasoning

Gunnar Misund
Bjørn Sigurd Johansen
Geir Hasle

SINTEF Informatics, P.O. Box 124, Blindern, N-0314 Oslo, Norway
{Gunnar.Misund, Bjorn.Johansen, Geir.Hasle}@si.sintef.no

Abstract:

In this paper we describe synergy effects of combining state-of-the-art Geographic Information Technology (GIT) with novel methods for planning and scheduling from the field of Constraint Reasoning (CR). We present a method for solving what we have called the Clear-Cut Scheduling Problem (CCSP), where the task is to assign clear-cutting times to regions in a given forest area over a long term horizon. The schedule must satisfy certain ecological, recreational and economical constraints, and, in addition, optimise on a number of partially conflicting criteria. Our approach is based on the combination of advanced spatial analysis and modern techniques for heuristic search. We have implemented a prototype Clear-Cut scheduling system called ECOPLAN. Empirical experiments have been carried out on a real-life test case consisting of a 500 stand forest property.

Keywords: *Forest Harvesting, Scheduling, GIS, Optimisation, Modern Heuristics, Tabu Search.*

Next: 1 Introduction

gunnar.misund@si.sintef.no

Lubos Mitas, Helena Mitasova, William M. Brown, Mark Astley

Interacting fields approach for evolving spatial phenomena: application to erosion simulation for optimized land use

We propose a computational framework and strategies for performing tasks necessary for evaluation and optimization of land use management within an advanced GIS modeling environment. Such tasks involve modeling of landscape processes, simulation of impact of human activities on these processes and optimization of preventive measures aimed at creating sustainable landscapes. A typical example of interaction between landscape processes and human activity is water erosion, a natural process which can be accelerated or greatly reduced by human intervention. In our implementation, the underlying 2D water and sediment transport equations are solved within approximate diffusive wave hydrologic and detachment/transport capacity models using stochastic methods (Monte Carlo). The methods employ distributed input parameters and enable us to simulate the impact of complex terrain, various soils and land cover changes on the spatial distribution of erosion and deposition. The presented approach together with land use optimization scenario is illustrated by an application to an experimental farm in Germany, carried out within 3D dynamic GIS environment.

This document requires a browser which supports tables.
The full size figures and animations can be retrieved by clicking on reference images.

INTRODUCTION

Efforts to balance the economic development with environmental protection have increased the demand for simulation tools enabling predictions of the human impact on landscape. In order to prevent irreversible changes and avoid costly, ineffective solutions the simulation tools should provide detailed spatial and temporal distributions of modeled phenomena. Statistical averages for the entire study areas or predictions only for a certain point, such as watershed outlet are often insufficient. New developments in GIS, especially support for multivariate temporal data processing, analysis, and visualization (Mitasova et al. 1995, GRASS4.2 libraries) make such simulations possible and GIS plays an important role in the development and applications of distributed landscape process models (e.g., Engel 1995(review with links), Vieux et al. 1996, Saghafian 1995, Leavesley et al. 1996).

The goal of our paper is to outline a methodological framework for distributed models based on the solution of the "first principles" master equations for multi-variate fields and use these tools for the distributed land use scenarios optimization. The basic platforms of our approach are described as follows:

- **Representation of phenomena as multivariate fields (i.e. as genuine distributed objects).** This advances lumped or semi-lumped description into the high resolution distributed one with advantages for spatial and temporal analysis. Such approach requires multivariate tools which deal efficiently with transformation from one discrete format to another, and with the possibilities of continuous representation for calculation of gradients, curvatures and other quantities. For these purposes we use the multivariate splines developed previously (Mitasova et al. 1995) and tested on a variety of 2D, 3D and 4D data (Table 1.).

- **Phenomena description and prediction based on solving master equations which determine the configuration or evolution of corresponding fields.** This enables us to perform simulations which are formulated in terms of fundamental physical processes such as flux, diffusion, etc., and we can use the known mathematical and physical apparatus to analyze the solutions. This level of rigor avoids vague concepts which often characterize various semi-empirical schemes and allows us to clearly specify inputs and governing parameters as well as understand the character of processes involved and final outputs. We can also build upon experience from other disciplines in using efficient and robust methods for solving the underlying master equations.

- **Formulation of cost functionals and their optimizations.** The proper quantification of objectives is very important for an efficient solving of desired land use practices. It is now well understood that the formulation of an appropriate cost or objective functional provides a powerful strategy how to deal with this task. The cost functional depends on the fields and includes also various conditions or restrictions. The human actions can change the character or properties of some of the input fields or modify their future evolution. The "space of human actions" is then explored for estimating the optimal solutions.

We illustrate these three fundamental concepts on a case of erosion prevention by optimized land use scenario. Water erosion, with its economical and ecological impact, is a typical example of problems targeted by our approach. It is a genuine space-time distributed phenomenon with several natural components such as terrain, soils, cover and climate effects. These can be naturally and represented by multivariate fields (terrain surface, terrain cover, water and sediment distributions, water and sediment fluxes, soil distribution, etc). The erosion processes can be described by continuity and momentum conservation equations. Some of these fields, especially the land cover, can be influenced or changed by human intervention and therefore will naturally enter the optimization process.

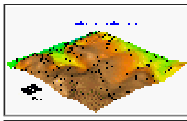
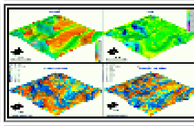
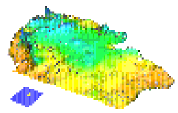
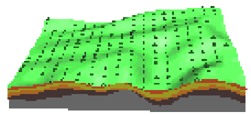
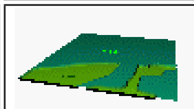
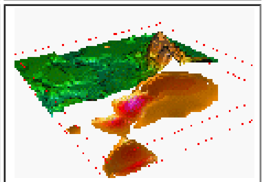
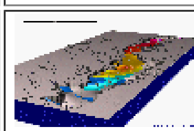
In the following sections we describe the **working examples of advanced methods and approaches** which enabled us to start from basic input data (terrain, cover, soils, etc) and get to the actual creative process of land use optimization in a fully distributed manner.

LANDSCAPE CHARACTERIZATION IN 3D SPACE AND TIME

The basic objects which enter into distributed models of landscape processes are given by functions which depend on the position in 3D space and time: *multivariate scalar and vector fields*. These fields represent various phenomena such as terrain, soil properties, land cover, fluxes of matter. They are usually represented in a GIS database in a discrete form as sets of points (sites, lines or polygons) or rasters. However, they can be transformed to continuous representations using expansions in an appropriate basis set such as multivariate regularized spline with tension (Mitasova et al. 1995), as illustrated by examples in Table 1. or by Hargrove et al. (1995). Such representation is not restricted to continuous fields, an example of effective handling of surfaces with faults using splines and GIS tools was developed by Cox et al. 1994. Phenomena represented by classes, such as vegetation or land use, can be represented as fields with faults as well (Table 1., land cover).

Table 1. Examples of landscape phenomena representations by multivariate fields
(*Click on the image to retrieve full size picture or animation*)

Representations of natural phenomena as multivariate fields modeled by bi-, tri-, and quad-variate splines		
phenomenon (field)	**3D (dynamic) "map"**	**GIS format (discrete representation)**

elevation		points (x,y,z), lines (contours), 2D raster (DEM)
elevation gradient and curvatures		2D rasters derived from z=f(x,y)
precipitation		points(x,y,z,t,p), 2D raster time series
soil horizons		points (x,y,z,w), 2D raster vertical series
land cover		polygon, 2D raster
underground concentrations of chemicals		points (x,y,z,t,w), 3D raster time series
concentration of chemicals in water		points (x,y,z,t,w), 3D raster time series

The fields are static or evolving in time. The evolution can be monitored and data from monitoring can be stored in a GIS database in various formats, for example, as time series of sites which can be further transformed to time series of 2D or 3D rasters (e.g., Table 1.,chem. concentrations). To predict the future states of these fields, we need to understand the processes controlling their evolution and formulate models simulating their fundamental behavior. The GIS software is being enhanced to support such simulations by providing the adequate data structures, including 2D and 3D floating point raster data (Waupotitsch and Shapiro 1995), multidimensional site data (McCauley 1995), support for temporal data (Brown and Shapiro 1995), methods for transformation between discrete and continuous representations using multivariate interpolation by radial basis functions (Mitasova et al. 1995), and tools for interactive multidimensional dynamic cartography (Brown et al. 1995).

FLUXES IN LANDSCAPE

The problem of landscape process simulations related to land use change have been studied by a variety of approaches, such as rule-based models, cellular automata, probability transitions models, (Berry et al. 1995) or models based on continuous fields and differential equations (Maidment 1996). We use the last approach, with continuous representation of fields and differential equations describing the fundamental conservation laws.

Water flux

One of the primary processes in landscape, influencing the distribution of soils, plants and people is water flow. While there are numerous empirical and process based models for modeling the water flux in streams and rivers, spatial distribution of water on complex hillslopes (crucial for soils and organisms) is still being modeled by rather rough estimates unable to provide sufficient detail for modeling of some important water related processes, such as erosion.

In general, the overland flow is described by the continuity equation:

$$\frac{\partial h(\mathbf{r},t)}{\partial t} = i(\mathbf{r},t) - \nabla \cdot \mathbf{q}(\mathbf{r},t) \qquad (1)$$

$$\mathbf{q}(\mathbf{r},t) = h(\mathbf{r},t)\mathbf{v}(\mathbf{r},t) \qquad (2)$$

where h is the water depth [m], i is the rainfall excess=rainfall-infiltration [m/s], $\mathbf{q}$ is the water flux [m.m/s], $\mathbf{v}$ is the flow velocity [m/s], $\mathbf{r}=(x,y)$ is the position [m], and t is the time [s]. The velocity $\mathbf{v}$ is related to h by Mannings or Chezy law and the momentum conservation equation (Haan 1994).

Current GISs provide the tools for the simplest approximate solution of the equation (1) for steady state flow and constant velocity, based on the upslope contributing area. There are numerous algorithms available for its estimation, such as D8 (Figure 1a), or an improved approach based on the vector-grid algorithm (Mitasova et al. 1995, Figure 1b). While these geometric approaches provide enough information for a wide range of applications (Moore et al. 1992), they become problematic when applied to areas with spatially variable cover and complex terrains with significant spatial variations in flow velocity.

A more realistic approximation which takes into account the flow velocity is based on the steady state solution of the equation (1) for the 2D kinematic wave (Figure 1c) and 2D diffusive wave (Figure 1d) approximation (Mitas and Mitasova, in preparation.). The 2D approximate diffusive wave solution can be found by solving the steady state form of the equation (1) written in the operator form as:

$$W[h^{5/3}(\mathbf{r})] = -\frac{e}{2}\nabla^2[h^{5/3}(\mathbf{r})] + \nabla \cdot [h(\mathbf{r})\mathbf{v}(\mathbf{r})] = i(\mathbf{r}) \qquad (3)$$

where e is the diffusion constant and W is the operator. We use a stochastic method to solve the equation (3). This approach is based on the representation of the solution h by a large set of random walkers (sampling points) which are propagated according to the Green's function corresponding to the inverse operator W. The diffusion term in (3) describes (approximately) a backwater effect and also helps to reduce the artificial features in water surfaces on hillslopes (Figures 1a,b,c) due to the flow-tracing on a regular discrete grid, which can cause problems for erosion/deposition modeling. This approach can be extended also for non-stationary event based modeling.

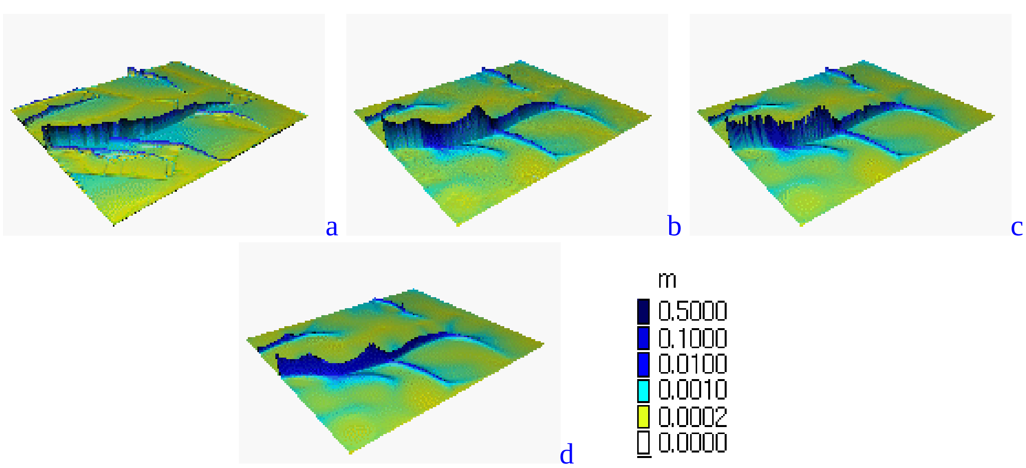


Figure 1. Steady state water depth estimated by a) upslope contributing area using D8 algorithm, b) upslope contributing area using vector-grid algorithm c) 2D kinematic wave approximation, d) 2D approximate diffusive wave

Other approaches which solve for temporal and spatial distribution of water depth during storm events are, for example, kinematic (Garrote and Brass 1995, Vieux et al. 1996) or diffusive wave models (Saghafian 1996), already integrated with GIS (r.water.fea, r.hydro.CASC2d). The choice of hydrology model complexity and realism depends on the type of application and for our current efforts steady state provides adequate information for assessing the impact of water flow on landscape at the time scale of days to years. The steady state solutions are also consistent with the new generation erosion model Water Erosion Prediction Project (WEPP) (Flanagan and Nearing 1995).

Sediment flux and erosion/deposition in complex terrain

Soil erosion involves detachment, transport and deposition. The interaction between soil detachment and sediment transport is controlled by water flux, terrain, soil and cover. This interaction is difficult to capture by traditional empirical models or models based on the geometrical analysis of terrain. While some of these models provide adequate tools for a qualitative assessment of erosion risk for large areas with complex terrain (Moore and Wilson 1992, Mitasova et al. 1996), they are insufficient for modeling of impact of spatially variable land use and simulation of erosion protection measures effectiveness.

The basic relationship for fundamental erosion processes is continuity of mass. For erosion by 2D overland flow, the continuity equation is (Foster and Meyer 1972, Govindaraju 1991)

$$\frac{\partial[r_s c(\mathbf{r},t) h(\mathbf{r},t)]}{\partial t} + \nabla \cdot \mathbf{q_s}(\mathbf{r},t) = \text{sources} - \text{sinks} \qquad (4)$$

where **q**_*s*=*r_s* . *c* . ***q*** is the sediment flux [kg/(ms)], *c* is the sediment concentration [particle/(m.m.m)], and *r_s* is the mass per sediment particle [kg/particle]. Further:

$$\text{sources} - \text{sinks} = D_r(\mathbf{r},t) + D_i(\mathbf{r},t) = C[T(\mathbf{r},t) - |\mathbf{q_s}(\mathbf{r},t)|] + D_i(\mathbf{r},t) \qquad (5)$$

$$T(\mathbf{r},t) = K_t[r_w g h(\mathbf{r},t) S(\mathbf{r}) - \tau_{cr}] \qquad (6)$$

where *T* is the sediment transport capacity [kg/(ms)], *D_r* is the rill erosion or deposition rate, *D_i* is the interrill contribution [kg/(m.m.s)], *C* is the first-order reaction coefficient dependent on soil and cover [1/m], *r_w* is the mass density of water [kg/m.m.m], *S(**r**)*=grad *z*(**r**) is the slope [m/m], *z(**r**)* is the elevation [m], *Kt* is the transport capacity coefficient, *g=9.81* is gravitational acceleration [m/s.s.s]. We assume that the critical sheer stress *tau_c* is negligibly small.

Rill detachment and deposition are proportional to the difference between transport capacity and sediment load (eq. 5). This relationship defining the interaction between sediment load and transport capacity (Foster and Meyer, 1972) is based on a stream power concept (Haan 1994) and can be expressed as:

$$D_r/D_c + q_s/T = 1 \qquad (7)$$

where *D_c=C.T=K_d.r_w.g.h.S* is the detachment capacity and *K_d* is the detachment capacity coefficient (rill erodibility).

We solve the continuity equation in 2D form for steady state water flux with small diffusive term with amplitude *d*, rewritten in the operator form as:

$$\mathcal{L}\varrho(\mathbf{r}) = -\frac{d}{2}\nabla^2\varrho(\mathbf{r}) + \nabla\cdot[\varrho(\mathbf{r})\mathbf{v}(\mathbf{r})] + C\varrho(\mathbf{r})|\mathbf{v}(\mathbf{r})| \qquad (8)$$

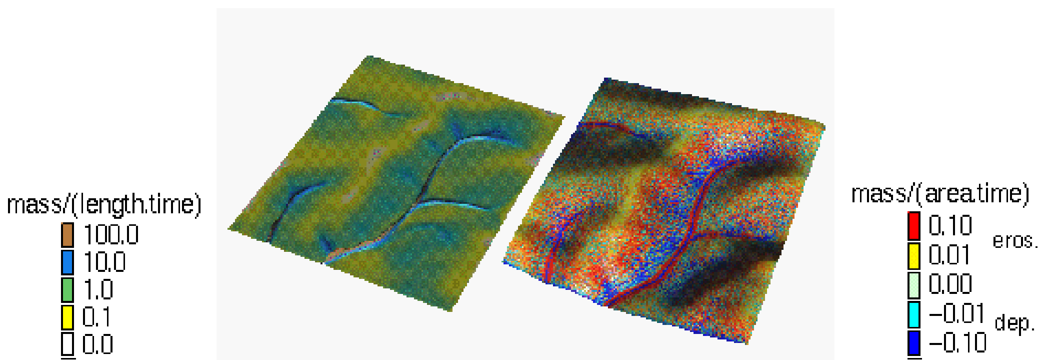
and then the erosion equation is

$$\mathcal{L}\varrho(\mathbf{r}) = CT(\mathbf{r}) + D_i(\mathbf{r}) \qquad (9)$$

where *rho=r_s.c.h*. The interpretation of the equation (8) is clear: the first term is the diffusion (which in our case is very small and represents the smoothing component of the soil transport), the second term is the drift driven by the water velocity **v(r)** and the third term is the 'potential' which is dependent on the velocity magnitude: the larger the velocity, the smaller the concentration of sediment.

Net erosion and deposition is then estimated as a divergence of sediment flux. Further details about this approach and comparisons with previous estimations of erosion/deposition by the directional derivative (Mitasova et al. 1995, Wilson and Moore 1992) will be given in Mitas and Mitasova, (in prep.).

The equation (8) is solved analogically as the equation for water flux using a stochastic method (Mitas and Mitasova, in prep.), illustrated by Movie 1.:



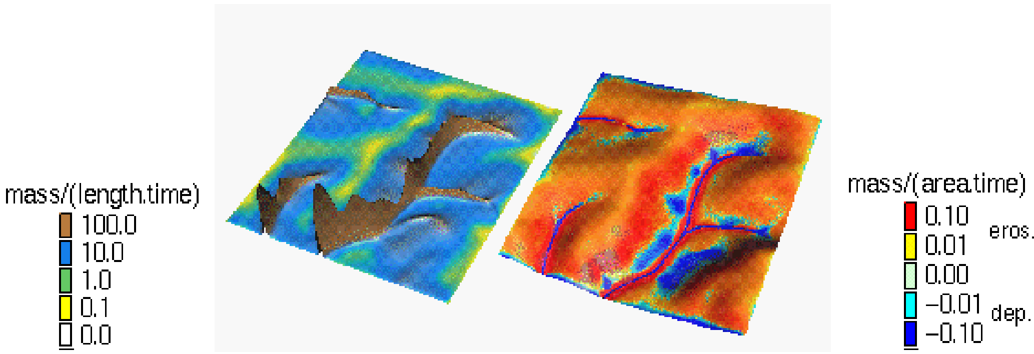
Movie 1. Solution of erosion equations by Monte Carlo, illustrated by a surface representing the sediment flux and by terrain with draped erosion/deposition

The animation shows that the approximate estimation of sediment flux is reached for a relatively small number of Monte Carlo sampling points. Accurate calculation of erosion/deposition estimated by derivatives of sediment flux, which are very sensitive to statistical noise, can be estimated to a given accuracy either by employing large number of walkers or by smoothing out the noise numerically.

Influence of uniform soil and cover parameters

In the erosion model the influence of soil and cover is represented by the following basic parameters: Mannings *n* , detachment rate coefficient (erodibility) *Kd*, and sediment transport coefficient *Kt*. These coefficients are functions of soil and cover properties such as soil texture, canopy, roots, management practices etc., and their estimation and development of various adjustment factors is described in WEPP (Flanagan and Nearing 1995). Constants for estimation of detachment and sediment transport capacity are still under development and detailed discussion of these parameters is beyond the scope of this paper. However, we will use the following examples to elucidate the role of these parameters in modeling spatial distribution of areas with erosion or deposition.

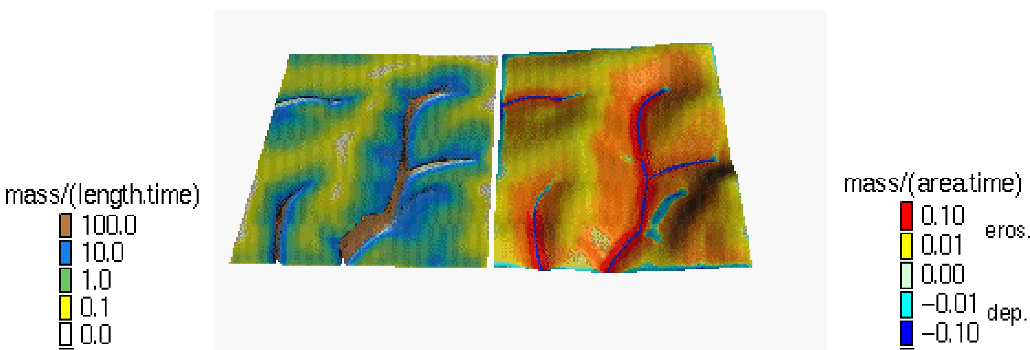
The most important parameter controlling the *borderline between the erosion and deposition* is the first-order reaction coefficient *C* related to the ratio of detachment capacity and sediment transport capacity. In the first example, we simulate the situation when the study area has constant transport capacity coefficient *Kt* but detachment capacity coefficient *Kd* increases, so that the ratio *C* increases from 0.0005 to 100 (Movie 2).



Movie 2. Change in the spatial distribution of sediment flux and erosion/deposition due to the change in *C* with increasing *Kd* and *Kt=const*.

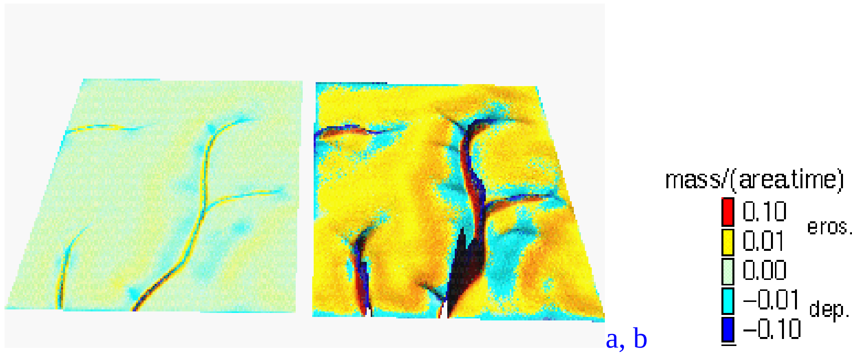
For small values of *C* (e.g., clay with very low fall velocities and low detachment capacity) water has the power to carry almost all detached sediment into the stream. For values *C*>1 (e.g., sandy soils with relatively high fall velocities which detach easily), the sediment flux quickly reaches the sediment transport capacity and deposition occurs relatively high in the hillslope. This is the case of transport limited erosion/deposition modeled by a simplified approach described e.g., by Mitasova et al. 1996 and Moore and Wilson 1992. For both cases the magnitude of sediment flux in the stream remains the same while the distribution of erosion and deposition over the landscape changes significantly. This simulation is a good example which shows that calibrating the erosion model using only the observed values of sediment flux at an outlet does not guarantee correct predictions of erosion/deposition on the complex hillslopes within the watershed.

Change in *Kt* while *Kd* is constant also changes the spatial distribution of erosion/deposition, however there is a big difference in the amount of sediment delivered to streams. For *Kt*<<*Kd* and *C*>>1 most of the detached material deposits before it enters the stream, for *Kt*>>*Kd* and *C*<<1, there is only a very small deposition and most of the detached material is delivered into streams (Movie 3). This example illustrates how the potential changes in soil properties and cover which increase transport capacity can trigger severe erosion.



Movie 3. Change in the spatial distribution of sediment flux and erosion/deposition due to the change in *C* with increasing *Kt* and *Kd*=const.

If *C*=const. and *Kt, Kd* change with the same rate, the spatial distribution of erosion/deposition is the same, and only its magnitude changes (Figure 2.)



a, b

Figure 2. Change in the magnitude of erosion/deposition with *C*=1 and increase in both *Kt and Kd*; a)n=0.1, Kd=Kt=0.0003 (grass on sandy soil); b)n=0.05, Kd=Kt=0.03 (bare sandy soil). Surface topography represents the sediment flux, color is the erosion/deposition.

It is important to note that the parameters *Kt and Kd* which are dependent on the soil and cover properties are interrelated and the change in one parameter is usually accompanied with the change in the second parameter. As we have demonstrated, it is their ratio, which plays important role in the spatial distribution of erosion and deposition. With better understanding of the physical basis of these parameters the analysis outlined above can be used for identification of those soil and cover properties which can be targeted for the most effective erosion prevention.

Erosion/deposition for spatio-temporal changes in soil and cover parameters

Erosion process is highly dynamic and its temporal variability can be modeled at various time scales from minutes (event based models such as ANSWERS or AGNPS), days (WEPP), to geological time (Moglen and Brass 1994). For land use management applications we adapt the concept used in WEPP and we simulate erosion under steady state flow for variable climatic, soil and cover parameters as they change during the year.

To evaluate the predictive capabilities of the proposed model, we have used elevation, soil and land use data from an experimental farm of the Technical University in Munchen, Germany (courtesy Dr. Karl Auerswald). We simulated sediment flux and erosion/deposition for the current land use with various cover and rainfall conditions. Comparison of the results with the observed spatial distribution of colluvial deposits (Figure 3) and with the pattern of linear erosion after an extreme storm (Figures 3, 4) indicates, that for this area with mostly sandy soils, the terrain controls the long term spatial pattern of deposition reflected in colluvial deposits observed in mostly concave areas (Figure 3a, Movie 4). Land cover seems to have more significant impact on the magnitude of erosion and short term linear erosion features (Figure 4).

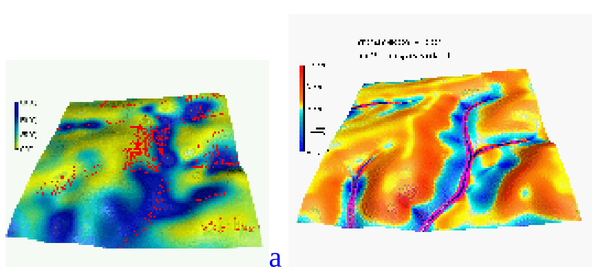

a b

Figure 3. Comparison of a) observed spatial distribution of colluvial deposits (depth in cm) and linear erosion features after the 150 year storm (red lines) with b) simulated erosion/deposition with homogeneous bare soil conditions.

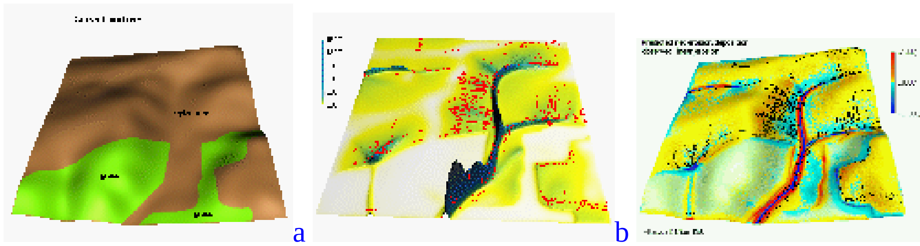
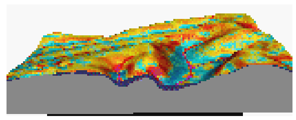
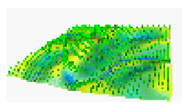


Figure 4. Simulated b) sediment flux and c) erosion/deposition with incorporation of the current land use influence.



Movie 4. Terrain and the observed depth of colluvial deposits, with predicted erosion/deposition draped as a color map.

We have also simulated the changes in erosion/deposition during a year due to the changes in rainfall and land cover, illustrated by Movie 5. and Figure 5.



Movie 5. Change in the land cover due to the plant growth and harvest

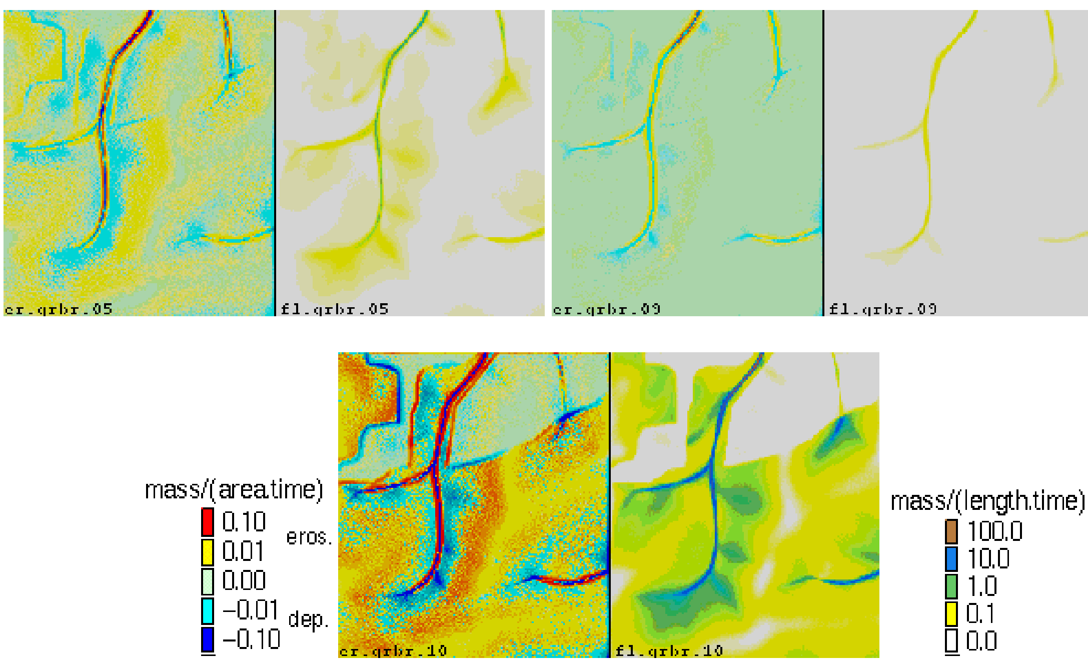


Figure 5. Impact of changes in rainfall and cover on erosion/deposition distribution and sediment flux in May, September and October.

The highest risk of erosion was predicted in October, when there is minimum cover and enough rainfall to produce significant runoff. Although May had the most intense storm, the erosion was lower due to the good cover provided by both the grass area and the agricultural area with winter wheat.

LANDUSE OPTIMIZATION

Human activity changes character or properties of landscape components (e.g. cover) which can be represented as an appropriate change in the corresponding field. These changes influence the natural phenomena through various interactions which can be described by the governing master equations. Well-defined and quantified impact of human actions on nature enable researchers to formulate objectives or costs in order to either predict the future development or, more importantly, to achieve a desired sustainable development. The desired objectives can include, e.g., maximization of land use for production or military training with minimized impact on environment, or prevention of unacceptable changes in environment in the given time horizon with minized costs (Johnston and Hopkins 1994). Because of the extremely complex nature of the problem, the optimization tasks are often out of the scope of ordinary techniques as they involve multivariate fields (possibly evolving in time) and also because of the special type of human action (like instantaneous point sources such as contamination which spreads out in the time horizon of a few years or clear cut of a forest with consequent erosion). Therefore this problem requires a formulation of general methods which can deal with complicated types of "configuration or state spaces". For our case we can define the *state space* as a set of fields (i.e., a particular set of multivariate functions) which describe components of the studied phenomenon. Available information and models such as initial fields values, are provided by a GIS and are used as inputs into the master equations and their solvers. In addition, we need to express the objective (cost) functional which is to be minimized within given constraints. The constraints can be formulated in the form of ``external" fixed influences (e.g. part of land cover which cannot be changed), thresholds on evolving fields (e.g. erosion beyond certain level is unacceptable) and so forth. The general form of the cost functional can be given as:

$$I = \int\int d\mathbf{r}dt F(\{z_i(\mathbf{r},t)\}) \qquad (10)$$

where I is the cost functional, z_i are the input spatio-temporal fields, F is a function which determines the cost for a particular set of z_i (point in the state space). In general, a minimization of (10) can be a very complex task. In order to carry out the minimization of the functional (10) we have to define the following:

- "distance" in the state space
- efficient representation of the fields which can be changed by the human impact (e.g. using appropriate basis function expansions)
- "movement" of the space of field configurations

Another important task is to formulate efficient minimization strategies. Because we are dealing with a multivariate problem which often involves non-linearities the cost functional can have many local minimas. This requires use of robust minimization methods such as simulated annealing or genetic algorithms.

A simple example of using GIS and the erosion model to optimize land use by finding a more effective spatial distribution of protective grass cover while keeping the ratio between the agricultural area and area protected by grass constant is illustrated in Figures 6 and 7. Under the current land use (Figure 6a), there is still a significant amount of sediment delivered to the stream (Figure 6b) with strong potential of creating rills and gullies (Figure 6c, dark red) which is in agreement with observations of big storm effects, presented in Figure 4. Redesigning the land use so that the protective grass cover is located in the highest erosion risk areas (Figure 7a) can dramatically reduce soil loss and sediment delivery to the streams (Figure 7b,c). The crest in sediment flux in areas with observed gullies disappears and is replaced by light deposition caused by the decrease in water velocity in the grass strip (Figure 7c).

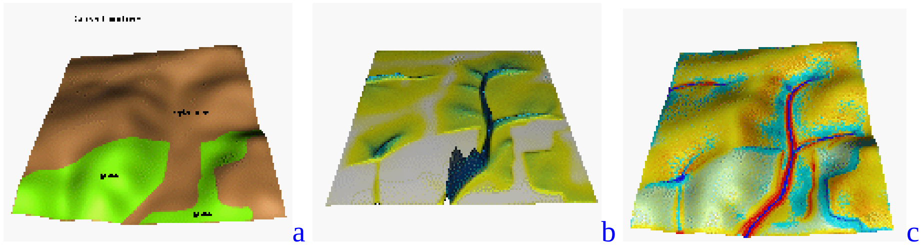


Figure 6. Current situation: a) land use, b) simulated sediment flux, c) erosion/deposition

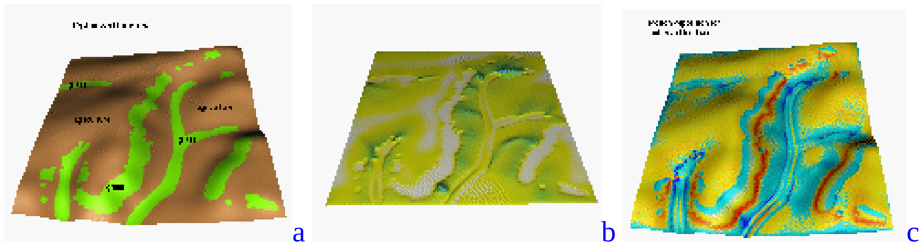


Figure 7. Situation with the grass area spatially distributed for more effective erosion protection : a) land use, b) simulated sediment flux, c) erosion/deposition

CONCLUSION

We have presented an approach to modeling of landscape processes in an advanced GIS environment which is based on the following developments :

- We have used multivariate regularized splines with tension for scattered data interpolations and continuous field representations for a multitude of tasks such as processing of input data, analysis and presentation of simulation results.
- We have developed and employed Monte Carlo methods for solving both water and sediment 2D transport problems. This approach proved to be robust and can be relatively easily generalized to include effects currently omitted (e.g., 3D infiltration process). The stochastic methods are also very well suited for distributed parallel computing.
- Extending the erosion model from traditional 1D water and sediment flow equations to fully 2D fields, supported by the GIS implementation, provided a new insight into the functioning of these models in a complex realistic landscape.
- Fully integrated visualization based on multiple dynamic surfaces helped in various stages of development, evaluation and applications of complex models, where interaction between the spatial fields is important.
- On the basis of these achievements we were able to produce land use scenario with a potential of significantly decreased erosion in a fully distributed manner.

Important computational components of our approach such as interpolation tool, water flux solver, sediment flux solver, scenario optimizer (under current development) are built as *functional units* with well-defined input, output and controls. Therefore these units can be used either as separate tools or within open GIS frameworks such as GRASS depending on the computational environments and size of tasks.

Using the presented application as an example, we believe that the GIS in future can become not only a powerful tool for providing and analyzing spatial information, but by extending its capabilities as a simulation and optimization tool, it can allow its users to find unexpected solutions of land management problems leading to practices which can be more effective at lower cost than the currently known conservation approaches.

Acknowledgments

This project was funded in part by the Strategic Environmental Research and Development Program (SERDP) and by the Army Research Office, Terrestrial Sciences Division. We greatly appreciate the continuing support from the USA Construction Engineering Research Laboratory (USA CERL) and the computational resources at the National Center for Supercomputing Applications (NCSA) in Urbana-Champaign, IL. We would like to express our gratitude to Dr. Steven Warren and Prof. Douglas Johnston for their support, and to Michael Shapiro, Olga and Roman Waupotitsch and Darrel McCauley for their contribution to the new GRASS development which made this work possible. Data used in this document were provided by Prof. K. Auerswald from Technical University Munchen, Germany, in cooperation with Steven Warren from USA CERL.

REFERENCES

Berry, M.W, Flamm R. O., Hazen B. C., MacIntyre, R.L.(1995). The Land-Use Change Analysis System (LUCAS) for Evaluating Landscape Management Decisions. *IEEE Comp. Science and Engn.*, (in press).
http://www.cs.utk.edu/~lucas/

Brown, W.M. and Shapiro, M. (1995). DateTime Library.
http://softail.cecer.army.mil/~brown/doc/dt_pman.html

Brown, W.M. and Astley M. (1995). NVIZ tutorial.
http://www.cecer.army.mil/grass/viz/nviz.tut.html

Cox, S. (1994). Toward a Fission Track Tectonic Image of Australia: Model based interpolation in the Snowy Mountains using a GIS.
http://www.ned.dem.csiro.au/AGCRC/papers/snowys/12agc.cox.html

Engel, B. (1995). Hydrology Models in GRASS.
http://soils.ecn.purdue.edu:80/~aggrass/models/hydrology.html

Flanagan, D.C. and Nearing, M. A. (1995). USDA-Water Erosion Prediction Project (WEPP), NSERL report No. 10 National Soil Erosion Lab., USDA ARS, Laffayette, IN

Foster, G.R. and Meyer, L.D. (1972). A closed-form erosion equation for upland areas. *Sedimentation: Symposium to Honor Prof. H.A.Einstein* (Shen, ed.), Chap. 12, pp.12-1-12.19. ColoradoState University, Ft. Collins, CO

Garrote, L. and Bras, R.L. (1993). Real-time modeling of river basin response using radar-generated rainfall maps and a distributed hydrologic database. Report No. 337. MIT, Cambridge, Massachusets.

Govindaraju, R.S. and Kavvas, M.L. (1991). Modeling the erosion process over steep slopes: approximate analytical solutions. *Journal of Hydrology* 127, 279-305.

Haan, C.T., Barfield, B.J., Hayes, J.C. (1994). *Design Hydrology and Sedimentology for Small Catchments*. Academic Press.

GRASS4.1 Reference Manual (1993). U.S.Army Corps of Engineers, Construction Engineering Research Laboratories, Champaign, Illinois, 422-425.

GRASS4.2 new libraries.
http://www.cecer.army.mil/grass/grass42_new_api.html

Hargrove (1995). Clinch River Environmental Restoration Program.
http://www.esd.ornl.gov/programs/CRERP/

Johnston, D.M. and Hopkins, L.D, TRAINER: a System for training requirements assessment and Integration with environmental resources.
http://ice.gis.uiuc.edu/trainer_html/trainer.html

Leavesley, G.H., Restrepo, P.J., Stannard, L.G., Frankoski, L.A. and Sautins, A.M. (1996). MMS: A Modeling Framework for Multidisciplinary Research and Operational Applications. *GIS and Environmental Modeling: Progress and Research Issues* (Goodchild, M.F., Steyaert, L.T. and Parks, B.O., eds.), GIS World, Inc.,pp.155-158.
http://www.terra.colostate.edu/projects/mms.html

Maidment, D.R. (1996). Environmental modeling within GIS. *GIS and Environmental Modeling: Progress and Research Issues* (Goodchild, M.F., Steyaert, L.T. and Parks, B.O., eds.), GIS World, Inc.,pp.315-324.

McCauley (1995). GRASS 4.2 Sites Format and API.
http://www.cecer.army.mil/grass/sites-api/index.html

H. Mitasova, L. Mitas, W.M. Brown, D.P. Gerdes, I. Kosinovsky and T. Baker (1995). Modeling spatially and temporally distributed phenomena: new methods and tools for GRASS GIS, *International Journal of Geographical Information Systems*, 9, 433-446.

Mitasova, H., Hofierka, J., Zlocha, M., and Iverson, R.L. (1996) Modeling topographic potential for erosion and deposition using GIS. *Int. Journal of Geographical Information Systems* (in press).

Moglen G.E. and Bras R.L. (1994). Simulation of observed topography using a physically-based basin evolution model. Report No. 340. MIT.

Moore, I.D., Turner, A.K., Wilson, J.P., Jensen, S.K., Band, L.E., (1992). GIS and Land-Surface-Subsurface Modeling. *Geographic Information Systems and Environmental Modeling*, (Goodchild, M.F., Parks, B., Steyaert, L.T., eds.), Oxford University Press, New York, pp. 196-230.

Moore, I.D., and Wilson, J.P., (1992). Length-slope factors for the Revised Universal Soil Loss Equation: Simplified method of estimation. *Journal of Soil and Water Conservation*, 423-428.

Mitas and Mitasova (1996). Distributed erosion/deposition simulations, in preparation.

Rewerts C.C. and Engel B.A., (1991). ANSWERS on GRASS: Integrating a watershed simulation with a GIS. *ASAE Paper No.91-2621*. American Society of Agricultural Engineers, St.Joseph, Missouri, 1-8.

Vieux, B.E., Farajalla, N.S. and Gaur N. (1996). Integrated GIS and distributed storm water runoff modeling. *GIS and Environmental Modeling: Progress and Research Issues* (Goodchild, M.F., Steyaert, L.T. and Parks, B.O., eds.), GIS World, Inc.,pp.199-205.

Saghafian, B., (1996). Implementation of a Distributed Hydrologic Model within GRASS. *GIS and Environmental Modeling: Progress and Research Issues* (Goodchild, M.F., Steyaert, L.T. and Parks, B.O., eds.), GIS World, Inc., pp.205-208.

Waupotitsch and Shapiro (1995). Floating-Point / NULL Values in GRASS Raster Maps.

http://softail.cecer.army.mil/~olga/fpnv.html

Lubos Mitas
National Center for Supercomputing Applications
University of Illinois, Urbana, IL 61801
lmitas@ncsa.uiuc.edu
Telephone: (217) 244-1971
FAX:

Helena Mitasova
Geographic Modeling and Systems Laboratory
Department of Geography
220 Davenport Hall
University of Illinois at U-C
Urbana, IL 61801
helena@gis.uiuc.edu
Telephone: (217) 333-4735
FAX: (217) 244-1785

William M. Brown
Geographic Modeling and Systems Laboratory
Department of Geography
220 Davenport Hall
University of Illinois at U-C
Urbana, IL 61801
brown@gis.uiuc.edu
Telephone: (217) 333-5077
FAX: (217) 244-1785

Mark Astley
Geographic Modeling and Systems Laboratory
Department of Geography
220 Davenport Hall
University of Illinois at U-C
Urbana, IL 61801
m-astle@uiuc.edu
Telephone: (217) 333-4764
FAX: (217) 244-1785

This paper, including color images and animations, can be also found at the following URL:
http://www.cecer.army.mil/grass/viz/SF.final/mitas.html

Nils FERRAND

Modelling and Supporting Multi-Actor Spatial Planning Using Multi-Agents Systems

Key-Words:

GIS, Spatial Planning, Multi-Agent Systems, Negotiation, Decision Support System, Multi-Criteria Spatial decision Aid, CSCW, Simulation of Societies

Abstract

Large and medium scale spatial planning is very complex, regarding either its object : the project and its environment (ecosystems, landscapes, socio-economics, etc.), or its process, which implies many actors, with different world's representation and related interests, and who are individually attached to a specific territory. Issues like exchange and coevolution of spatial representations between many distant actors, negotiation support and simulation, multi-actor multi-criteria decision support for continuous spatial planning, are usually not addressed by current GIS. We propose to use Multi-Agents Systems (MAS) to enhance or develop such functionalities. We present two approaches: in the first we use Multi-REACTIVE-Agents Systems to solve the complex spatial optimization problems encountered in the search for least environmental impact area for infrastructures, where environmental sensitivities, structural constraints and different localized actors decision systems are used. In the second approach, we use Multi-COGNITIVE-Agents Systems to support and simulate the exchange and dynamics of spatial representations and policies, considering the general political values, the specific spatial constraints, and the socio-relational characteristics of embedded actors. We conclude on the general interest of using MAS for spatial modeling, and argue ab out the development of specific MAS based GIS.

Introduction

As we are more interested in real world problems than in computer issues, we will address here a social and political process which is extremely important for the organization of the territory: spatial planning in its multi-actor dimension. We propose an overview of this process, which will show to be very dependent on the representation of space by the embedded actors. So our modelling deals with how world is perceived and communicated, and not how it is really. This perspective seems to be more operative in terms of decision making. But it is opposed to usual GIS principles.

In order to implement related information systems, we propose to use Multi-Agent Systems. We describe them quickly, and then present three applications: the first for spatial project analysis, the second for spatial negotiation support, and the last for spatial negotiation simulation.

Multi-Actor Spatial Planning: an Overview

Spatial planning is nowadays an extremely sensitive issue, especially in Europe, where old traditional settlements, high population densities within some critical areas, and complex political influences and relations, lead to a situation of permanent crisis regarding the use and destiny of lands. Environmental awareness since the mid 60's has brought new issues and related procedures, which tend to widen the potentially embedded system of people and constraints. Consequently, the process of spatial planning is hardly controlled, highly decentralized, and for wide area and major projects, can last for decades. Furthermore, the rationality of final solutions is not always obvious, regarding either social or environmental aspects.

So, being stated the importance and difficulty of these processes, it justifies the many studies and developments of dedicated information systems which were made. But among those, only few address the whole problem, including environmental assessment, spatial multi-criteria decision making, expert practices, multiple parties, consensus query, and negotiation. So, in this first part , we will try to analyze and describe precisely the processes to be supported. That way, we expect to be able to specify and develop dedicated information systems which could fulfill the different needs in an integrative way. Our reference here will be the french case, which can be somehow different from some countries, because of the extremely distributive (conflictuous, rather than cooperative) nature of the process.

General overview of spatial planning: a social construction of representations

Spatial Planning is essentially a *decision* process, which means that it includes at least the four classical (Simon, 1977) stages : intelligence (information query), design (of solutions), choice, and review (execution and feedback). But it has also other features :

- **Multiple scales**: in space (from the close neighborhood, to the nation, or even continent), in time (from an instant for the perceived feelings of negotiators, to the centuries for the evolution of large scale ecosystems), and in organizations (from the one-to-one interaction, to the transnational organizations);
- **Multiple actors:** except in some very local and specific situations (like a landowner planning some limited works on its own territory, with no visual or ecological impact), spatial planning implies many actors, with different interests, social and organizational position, spatial attachment, and personal qualities;
- **Multiple objectives**: environmental, political, personal, economical, etc.
- **Multiple modes** : integrative (toward consensus through cooperation) or distributive (toward compromise through negotiation);
- **Multiple criteria**: ecology, economy, landscape, agriculture, laws and regulations, culture, society, ...

So, three main systems are concerned: the *environment* (the territory and related elements), the *project* itself, and the *set of embedded actors*. Now, it appears that practically none of this real systems is operative; because as part of a decision process, ***only the representations of the environment, the project, and the society, by the effective decision makers (those participating in the applied decision), can induce reality at the end***. This assumption relies widely on the proposals of the constructivist school of Palo-Alto (Watzlawick, 1978). It implies that **spatial planning is totally dependent on its social and relational context**. So it is almost impossible to modelize spatial planning without a model of social systems and its related decision making process. Furthermore it should be considered that spatial planning is in fact a co-evolution or a co-construction of shared and personal representations of space, which converge through the game of interpersonal relations. Thus, even if basically representations are built on top of mediated reality, it is fundamental to remark that in the perspective of modelling and supporting spatial planning, we will have to assess how spatial representations are built or exchanged and evolve. Such knowledge could also help to promote the cognitive ergonomy of our systems. Finally this conception of spatial planning as a social construction of reality (the territory), emphasizes **the contingency of information**, and bring the relativity of information as a constant background. To summarize : *there is no data in spatial planning, only models*.

In the following, we will distinguish three levels : the first deals with the space itself, and the state of the environment; it consists in analyzing the potential impact of the project, and proposing alternatives. The second level is linked to the social and political context; and address the nature, and the construction of representations of the space and the project. The third level is reached when representations are exchanged, and when negotiation occurs.

Environmental Assessment by Experts

As an instance of the "intelligence" stage of Simon, the first part of the planning process consists in making what is usually called "Environmental Impact Assessment" (EIA, or EIE in french). Such work is realized by experts, chosen after their proficiency in the different domains to be considered (and the price they propose...) They are independent people, specialized in this activity. Their role is mostly to commit themselves into the conclusion they give. In fact the technical or scientific content of their conclusions is usually not at stake. And finally the decision makers, who are politicians, or administratives, just use experts as producers of "packaged solutions" to be compared. In the regular methodology we refer to, their specific work can be divided in three subphases (after the request by initial demander):

1. **Delimitation of the Studied Area** : choice and justification of the territory which will be considered; validation by requiring actors.
2. **Analysis and cartography of "stakes"**: intrinsic values attached to the area, for all the different criteria, and independently from the project itself. It gives a "state of the environment" before the project, and is used in the next phase. This phase relies on local investigations, enquiries, bibliography, and data collecting.
3. **Analysis and cartography of sensitivity**: combination of the stakes (values) with the nature of the project. It gives for each point of the space and for each criteria the potential loss of value (quality) if the project would go there. This classification is usually quite simple and with a finite discrete scale. But it requires most of the

expertise to be able to assess the potential impact at different scales and for all the criteria. The commitment of the expert is mainly here.

Next is a phase of integration of the different maps of sensitivity in order to propose a set of alternatives, with attached information about each of them. But, in a clean decision process, this last phase can't be done by the expert, as it includes a mixing of different criteria, which assumes in some way that a weighting is chosen. And such ranking of criteria is of the political domain. So this phase of "design" is no more directly related to the environment and the project; it is already a part of a social and political process.

Social and Political Issues

As we stated before, spatial planning is a multi-actor process which cannot be separated from its social context, and which consists in the construction and exchange of representations of the space and project between the actors. So, the first question is to determine what representation (model or image) is used by each actor. Our assumption here is that the results of the expert analysis of the environment, and the proposed mono-criteria maps of sensitivity, won't be discussed as such. If so, then the topic is moved from the decision field to the scientific or technical controversy, which is not our subject here. Then we consider that the correct "model of representation" is a complex map, valid only for one actor, and giving in one hand the possible alternatives as thought by this actor, and in the other hand the perceived global (in terms of criteria) value attached to the different places in the area.

Actors can be extremely different, regarding to their interests, values, relational abilities, or spatial attachments. But in a first step, we will only consider differentiation of actors after their "political position", or value system. To model this we use the classical approach of multi-criteria decision theory (Vincke, 1989), which consists in weighting the different criteria. So, to have an idea of different representations or "points of view", we need to combine those weightings and the sensitivity maps as given by experts. We could get that way a spatial system of preferences attached to a certain type of actor.

To refine this, in a second step, we should include a tolerance system above the weightings of criteria. It would describe the acceptance by actors of a modification of their position regarding a topic.

Now, in a third step, the spatial situation of actors is to be taken into account. Obviously the political position of actors is not homogeneous in space. Some are officially responsible for a specific area, some have personal interest here or there (house...), etc. So, rather than a uniform system of weighting, we can use a mapping of weightings attached to each actor. In a way, such a map is a model of the spatialized system of decision of the actors.

The approach proposed here address only one single actor representation of space and project. But in the context of multi-actors spatial planning, the collective spatial representation of a group is more operative. So we need not only the collection of individual decision (political) systems, but also their combination. If we have a set of embedded actors, the collective representation will result from the collective decision system which reflects the exchange of power position between actors. This model is more difficult to build, and for a sake of simplification, we consider that each actor as a certain voting power for each area, and that by combination on each point of the votes coupled to individual weightings, we get the global collective decision map. This approach doesn't consider any type of negotiation at this level; it only deals with the fact that actors are spatialized, and that they are organized with a basic system of power. So the representation we get here is a rationalized image, eventually different from the real situation, which includes external constraints (brought by actors in a negotiation), and relational effects linked to the context itself of collective decision making. To conclude here, one can remark that we need an inventory of potential actors, and a mapping of their location. Such information is difficult to get, and some assumptions are usually made.

Negotiation

Given the incomes of the expert about the state of the environment, and some assessment of the decision system of embedded actors, with its collective counterpart, we can address the fundamental phase of negotiation. When actors meet to negotiate a planning project, they make proposals, exchange points of view, and argue based on spatial or overall issues. The evolution of the negotiation depends on what is said, how it is said, by who, to whom, in which context, etc. It's not the object here to dissert about negotiation, but it should be noticed that most of the arguments exchanged during planning negotiations are either directly of a spatial nature (location, extension, topology), or at least georeferenced. Furthermore the type of messages that are sent is such that most of their content can be mis-interpreted by the receiver, because of a lack of common knowledge or common representation of the space. Finally, following the assumptions made above, the objective of negotiation is to make actors exchange representations and let their own evolve until an agreement is found, which means that either a common spatial solution is found (consensus), or some actors are obliged to give back some of their pretentions, and will get or not some counterpart. Whatever is the perspective about negotiation, actors will have to exchange geoinformations.

Required and Present Information Systems

After a brief description of the process we consider, we can propose a functional analysis of the related needs, and observe quickly how present technical approach fulfill the requirements. As a consequence, we propose some alternative technical answers.

Related Needs

As a principial for our whole work, we set that i*nformation processing systems should be developed according to established needs from the practice, and not as a proposed usage of already common systems or softwares*. In our context, as we stated above that most of the information used in multi-actors spatial planning is models (representations), rather than data, then we should emphasize the dynamic and interactive construction and analysis of spatial models. The background idea can be expressed by: **give the actor some bricks to build his world, and not a world to fit in.**

What are the needs ? We give here a functional classification, which we will refer to later:

- **Information** : retrieval, organization and presentation of the information, coming either

from data bases or specific collection on ground.
- **Modelling**: construction of models (symbolic translations of particular representation of the reality), which provide new informations, and can structure gross data.
- **Problem Solving**: finding a solution to some precise problem in space, as described by its data and some constraints on the solution (including optimization, and some other artificial intelligence approaches).
- **Multi-Criteria Decision Aid**: supporting the choice among a set of alternatives described by values for criteria.
- **Negotiation**: all relational process occurring when information is exchanged during collective planning.
- **Communication / Publishing**: mediating and exchanging information, and related technical support and form.

The different functional elements should be integrated in a coherent environment, dedicated to unskilled users.

Present Technical Answer

Presently, each of the functions is addressed by the following domains or technologies (refer to the glossary for abbreviations):

Information	DBMS , GIS , Networking
Modeling	GIS , KBS , DSS , MBMS , Simulators
Problem Solving	Optimization, AI , MAS
MultiCriteria Decision Aid	MCDA : ELECTRE, PREFCALC, PRIAM,etc.
Negotiation	NSS , CSCW , Simulators
Communication / Publishing	CAD , MultiMedia, GIS , Virtual Reality, NetWorking

Proposal

In our project, we will instrument only two functions: problem solving and negotiation. Because we estimate that the others are already filled by commercial products, whereas there is no answer for the complex integration of multiple spatialized criteria with decision systems distributed through space, and regarding negotiation, the principle of collective construction of spatial representation is not yet used.

We present thereafter the Multi-Agent approach and argue about its efficiency for our context.

Multi-Agents Systems

The following is abstracted from a general presentation for spatial issues in (Ferrand, 1995).

Origin and Motivation

The development of MAS followed two directions which fit the present main families in DAI . The first proposed to distribute problem solving between different modules, called agents, in order to distinguish different expertizes, to reuse common systems already developed, to mix heterogeneous processes, to gain openness and to get 'emerging functionalities' from the magma of interactions. It went through multi-expert systems with various organizations, and led to the present cognitive agents systems. This thread is mainly related to the symbolic approach, and it has shown its utility for solving complex problems. Cognitive agents have an explicit representation of their environment and of other agents. They have a memory; they build plans, selfish or cooperative; they exchange complex messages; and usually only few of them are used together in a system.

The second thread is more recent and can be called the reactive one. It has followed from many observations about information processing in nature, where 'intelligence emerge' from the interactions of multiple simple entities, acting on the base of direct reaction to stimuli. From this point of view, reactive agent systems are close to the connexionist paradigm, and many of its issues are shared. But the reactive paradigm also rely widely on ethology, where it is observed that extremely structured objects or processes are built by colonies ('swarms') of very simple animals, without intelligence, like ants or bees. In that perspective it is close to Artificial Life.

General definition

In our conceptual definition, *a multi-agents system is a set of agents interacting in a common environment, where an agent is an entity living in this environment and able to modify both its environment (communication, decision, action) and itself (perception, reasoning, learning). The environment is composed of all the entities from the addressed universe which are not agents.* So an agent can perceive and represent its environment, it can communicate with others, and it has an autonomous behaviour depending on its observations, knowledge and interactions. 'Autonomous' means mainly here that its behaviour is attached to the agent (self-control) so that whatever the social or environmental context is, it will have the same functional response to stimuli. Therefore no external control exists and impose an activity referring to informations and behavioral laws unknown by the agent. It should be noticed that such an agent can recursively be itself a system of sub-agents.

In the cognitive case, there is an explicit and symbolic representation of the environment and other agents. Cognitive agents handle goals, plans and resources. They have an expertise expressed in symbolic terms. Their communication with others is managed explicitly and the exchange of messages follows some fixed protocols. The reference domain is more or less sociology.

In the reactive case, there is no explicit representation of the environment or other agents, no memory, and the behaviour of agents is based on sensory-motor functions. Such agents are in fact complex automata with the ability to control their interaction pattern. One big interest of the reactive approach is that with such a simple behaviour, one can expect that some formal results about the global dynamic of the system could be obtained.

The implementation relies on the design and building of code modules that fit each agent.

Present usual approach consists in using "Object Oriented Languages" and dedicated environment or class libraries, but this choice is up to the designer. The main issue in the implementation is the management of parallelism. It can be real parallelism, if possible, but more often it is simulated parallelism, either with UNIX processes (using PVM for instance), or simply with regular synchronous or asynchronous iterations.

Application to geographical issues

There are two possible use of Multi-Agent Systems. The first one is simulation, where a computer model of the real world is built and used as a laboratory to study some process. The second approach is problem solving. It deals with the same issues than spatial optimization, which is a very classical problem in AI. To summarize, it tries to find a shape or pattern in space and time which fits some constraints.

Simulation of Spatial Processes

A Multi-Agent based Simulation consists in mapping the agents onto the modeled objects of the observed system. The closure, the interactions, and the initial state are also identical. Agents' state is defined by characteristic attributes, including at least its position, but also all other changing parameters. In the classical example of ecological simulation, agents would be all the modelled beings, with their position, species, age, sex, size, health, etc, as attributes, and their feeding, reproduction, moving mechanisms as processes. The environment would be the hypsometric model, the vegetables, the weather, seasons, etc... Many agents can be created and let evolve in their environment.

Spatial Problem Solving

As stated above, the issues are the same than those of classical optimization. But there, an utility or 'energy' function is defined for the pattern, which comply with the constraints, and a solution is found using minimizing methods . Whereas in the MAS paradigm, the pattern is divided into agents which will interact with the data (their environment) and with the other agents. As soon as the problem is concerned with two dimensional patterns (lines, zoning, networks), this approach is very relevant. The division of the pattern in agents leads to a set of subproblems localized in space, which can be autonomously solved. After their construction, agents evolve autonomously, and in parallel. If it doesn't stabilize, it means that no solution was found. Then the constraints have to be relaxed, or some parameters, like the agent sampling of the pattern, must be changed. If it stabilizes, than the reached pattern is a solution as it means that each agent (hence each part of the pattern) fulfills its requirements both versus the data and versus the pattern. If another solution is required, the system has to be reinitialized in a distant state. In other words, any steady state is the expression of a solution to the problem. This approach allows several variations: the problem can evolve with time; it can depend on the position; different problems can be combined easily; the stability of the solution can be analyzed just by changing some parameters and letting the system reevolve; qualitative and uncertain constraints can be used.

Multi-Agent Based Spatial Modeling

The use of MAS supposes to come back to a naturalist approach, because a description of the objects and processes is needed. But a qualitative analysis is possible as the computer model support such processing. In particular, symbolic rules can be used. According to the experiences we have made in relation to expert modelers (Ferrand and Demazeau, 1994), the type of modelling we require is very simple and natural to skilled experts.

We can accept sparse information as soon as they are connected in terms of processes. Furthermore, any type of information can be integrated : rules, functions, differential equations... And finally the model is entirely open: it is possible to add agents of any type at a low cost.

The modelling procedure is a classical systemic approach:

1. choosing the active objects of the model,
2. closing accordingly the system and defining the environment
3. defining the relations of objects with others and with the environment,
4. defining the actions of objects,
5. defining the effect of actions

Using a Multi-Agent implementation, it becomes operative.

Multi-Agents Based Systems for Multi-Actor Spatial Planning

We present thereafter three different systems, based on the Multi-Agents approach, which are applied to Multi-Actor Spatial Planning. The first is dedicated to spatial project analysis , where multiple criteria are used, and multiple actors with different points of view are embedded. The second is a project which deals with the support to negotiation of spatial projects and exchange of geographical informations and other type of arguments between different actors meeting on electronic networks. We introduce a specific agent based approach by implementing a mediator agent concept. The last is a tentative simulation of spatial negotiation , considering environmental constraints, political choices of actors, relational effects, and contextual dynamics.

A Support System for Spatial Project Analysis : SMAALA

After a long time spent in observing the practice of environmental assessment experts, and with a common analysis, different problems appeared which were mainly related to the phase of design , consisting in integrating environmental sensitivity maps into a proposal of alternative solutions:

- high cost and difficulty to reproduce the synthesis map for different demands
- no clean separation of expert input and political decision effects
- limited use of the synthesis map as a negotiation support
- no possibility to build a map for each possible point of view

The implicit request was to build a system which could integrate the sensitivity maps given by the experts, the structural constraints of the project, and a political position (an actor description) expressed by a weighting of the criteria, into a map of least impact solution for the project. Further extensions include spatialization of decision systems, sensitivity analysis of different solutions, and proposals for consensual alternatives.

Our answer was the SMAALA system (Systeme Multi-Agents d'Aide a la Localisation d'Amenagements, or Multi-Agent System for Support to Infrastructure Localization), initially developed on PC, and now transferred to UNIX SUN stations and Macintosh.

The application field retained for the first version of SMAALA is the planning of linear infrastructures (roads, railways, electric lines) through territories. Formally, such planning is mainly an optimization like problem, but multi-dimensional and with spatialized constraints and decision system (similar to spatialized optimization process). Because of a structural similarity of the problem with some former experience of complex image analysis, and robotics path planning, we decided to use Multi-Reactive-Agents Systems (MRAS) to solve it. The choice was motivated both by the accuracy of MRAS for complex spatial problem solving, because of the structural analogy of the problem and its processing (Ferrand, 1995), and by the fitness of the MRAS model to the mental declarative models used by the experts.

SMAALA was developed in 6 phases:

1. Multi-Criteria Spatial Decision Aid

The first level of the system is the pure least impact path finding, for one decision system, uniformous on space. SMAALA accepts the different sensitivity maps given by the experts, one for each criteria. It takes also as an input the morphological constraints of the linear project (curve, slope), and its topological constraints: start and arrival points, off-limit area, etc. Finally it uses a weighting of the criteria, simply given as an n-dimensional simplexes, and which describe then decision system of one actor. According to some process parameters which set the precision, SMAALA gives the different solutions for this actor, as a set of path, following the constraints, and minimizing the overall impact according to the relative importance of criteria given by the weighting. Any type of further spatial analysis (length of the path, statistics of impact) can be processed with classical GIS tools.

The principle of the optimization can be explained quite simply by the following metaphor: let assume that we plan an electric line between two points, and that each pylon is active and can move. It has perception ability toward its environment, and it is linked with other pylons by the line. They constitute the set of reactive agents. The problem is to find a position for each pylon so that the global impact of the line is minimized and the structural constraints verified. Then the pylons / agents individually will solve a very simple local problem which is: 'where should I go in my close neighborhood to decrease my impact ?' And then: 'Ok, but can I go there considering what my neighbors do ?'. So starting from a random choice (or straight solution) we let the system evolve (synchronously) till it gets satisfaction. To compute the local decision, we use the weighting as a parameter, and accept a direct normalized aggregation (utility based local decision making). By iterating different initial position, we finally get the collection of attractors of the system which show the solutions. Solutions can be given with a width showing the relative size of attractors (area of equal impact).

One can remark that regular Multi-Criteria methods (utility based, or overranking) cannot be used because of the continuity of the solution space. However, at this step, usual optimization (optimal path finding) is also possible, as the decision system is homogeneous and isotropic. But with the MRAS approach we allow the use of parallel processes directly, and facilitate the analysis of the proposed solution through dynamic modification of the parameters (including weightings) with fast reconfiguration of the solution.

To summarize, this level gives the representation of the solution to a planning project for one actor who considers all the space the same way. It should be noticed that the system usually gives only a set of possible path, and not the best one. The final choice remains to the decision maker (eventually with usual tools like ELECTRE).

2. Spatial Decision System

The second level includes the fact that decision systems of actors are spatialized: for instance the major of a town as not the same point of view (political position) in its town, and outside, and if he has other interests in some area, he can have still another position there. Then, instead of using a single set of weightings for the criteria, we replace it with a map of the position of an actor. For each point in space, there is a different weighting. Furthermore, it is established by experts that the least impact path finding process is anisotropic: the impact of an electric line following the main direction of a valley is reduced regarding to the same crossing the valley. Such parameter is implicitly integrated by human experts during his assessment. So the local aggregation result also depends on the local direction of the linear solution.

To implement all those constraints, we ask that the agents (pylons in our metaphor) use a spatialized and contingent local decision system: a different type of optimization for each point in space, and each direction of the current state of the solution. We suspect that such constraint is difficult to handle through usual analytical optimization process.

3. Sensitivity of a solution

The third level address the sensitivity of the found solution to a change in the actor's position: what happens in term of representation of the project by the actor if he changes his mind regarding general balance of criteria, or otherwise, what is the acceptability of modification of his proposal ? This question is very important because it prepares the ultimate negotiation by assessing the "open doors".

It is quite simple to get this result. In a first step we get the solution with a first system, and record it. At this stage, agents are localized in space as a sample solution. If we perturbate the weightings, agents will adapt automatically their position to fit the new constraints, and then give a new solution to compare with the first. The only difficulty is to choose a perturbation process in the n-dimensional criteria space.

4. Least Risky Solution

Having an image of the preferred solutions for each actor, we can compute the least conflictuous (least risky) solutions. Such solutions are the least sensitive in terms of decision system. Their computation is expensive, because the process consists in sampling the possible decision system with a common step (building a grid of values of criteria) and applying them to the SMAALA regular finding. Then for each type of initial weighting, corresponding to every possible type of actor, a set of solutions is found. By iterating this process, and selecting each time the least sensitive solutions (with a threshold and a limited number of paths), we get a set of expected least conflictuous solutions. It is useful when embedded actors are unknown before the negotiation.

In practice the number of observed solutions is limited, whatever is the context.

5. Compromise Solution

A rational compromise solution can be sought when the actors are known previously. All the solutions (representations by actors) are computed, and by superimposing maps we directly get the common paths, if there are. At least some parts are common, and can be used as a start for discussion.

6. Multi-Actor Solution

This last level adds a relational dimension, which is more subject to discussion. The idea is to take into account the fact that actors have different power and voting position. We have different possibilities: put a weighting on the global decision system of actors, give a priority to some actors for specific criteria (notion of expert), or even spatialize the relative importance of actors, each one having its own "territory". All those approaches are possible, and suppose to require that agents (pylons) use a pre-computed local system of decision to integrate the different environmental values. We obtain the collective solution that is potentially found, without any consideration of relational effects, or contextual negotiation processes. It assumes that enhanced actors knowledge is available.

Spatial Negotiation Support in a Multi-Agent Approach

This second application belongs more or less to the Computer Supported Cooperative Work (CSCW) thread. Its objective is to support negotiation of spatial planning between distant actors using electronic networks to communicate. It is a current project (Koning & al., 1995) developed within MAGMA group as an instance of its specific Multi-Agent approach, and implemented mostly in the JAVA language from SUN . There are two main aspects: the first is the possibility of collective work with shared geographic informations, exchanged through networks like Internet. We won't address it here, as there are already strong related programs (OpenGIS , GEOMED). The second aspect is the possibility to facilitate and "mediate" negotiation about spatial issues. In this prospect, a proposal was made, based on Cognitive Agent Systems. The main principle is to attach to each actor an agent "assistant" or "mediator", who knows its actor and the other assistants, and tries to help the process. It is not the object here to address the pure CSCW aspects. So we focus only on the geographical aspects of the mediator. Its main related duty is to handle georeferenced arguments in order to control and illustrate the spatial organization of issues. For instance mediator should provide the actor with any type of information which would be spatially related to a presently evaluated proposal. If a road is at stake somewhere, the mediator should give its actor info about any close object, or look for regulations about neighborhood of roads, or enquire about possibly interested actors. Meanwhile then assistant should also check the spatial coherency of arguments. For instance if a road is not to be less than 100 meters to a river, it should check that no argument break this rule.

To precise the function and content of the assistant we distinguish the following parts in the cognitive agent architecture:

- **Interface:** presentation of the state of the negotiation process. It should propose some maps of the different proposals, and illustrate the spatial reference of the exchanged arguments. It should also permit the access to informations about other actors (position, location), and other spatial data-bases.
- **Reasoning:** In terms of spatial issues, it consists mostly in assessing the topological relations of objects and arguments, and permanently checking the validity of constraints. This part uses the knowledge base for data.
- **Knowledge Base**: 3 sub-parts are included: representation of the world including the environment and the project, representation of actors including "its" actor and others, with data about their spatial interest and position, and mediating laws, which state how the negotiation is led, and in particular what is the spatial extent of issues (how far from a project should actor be acknowledged).
- **Communication part**: handles all the telematics issues, and can use the Knowledge Base to locate some actors.

This architecture is quite a classical Multi-Agent one, and is implemented using a regular platform containing packages for Agents, Environments, Interactions and Organizations. But in this first stage, we don't stress the presentation and interface aspects. So, maps are still not used as a medium, and we are expecting JAVA applications handling the exchange and manipulation of maps.

Simulation of Spatial Negotiation

As we are interested in supporting multi-actor spatial planning, we would like to be able to prepare negotiations and foreplan different strategies. In that perspective, and as there is no strictly rational approach, we try to develop a simulator of spatial negotiation, which could be used as a virtual laboratory. Negotiation as a whole is already much studied (Raiffa, 1982 ; Dupont, 1990), but spatial project negotiation is not usually addressed, as it much more difficult to discuss about a continuous space of solutions, than among a finite set of alternatives (as in usual business negotiations). Furthermore, in the context of multi-actor spatial planning, the closure of the system is extremely far. Many actors can demonstrate their interest and appear suddenly in the process. So it is mainly an open negotiation, very difficult to limitate. And the behaviors of actors are very complex, with different roles and hierarchical levels.

The analysis leads to two levels: the first is conceptual and address the modeling of actors as individuals, the second is instrumental and address the implementation of such a model into an effective simulator.

The model of planning actor we propose includes:

- **A value system**, giving the absolute political position of the actor,
- **A constraints system**, giving his relative set of contemporary interests and obligations,
- **A relational pattern**, describing his social ability to communicate and convince, and his social interests.

Two dynamic perspectives are available: in the first a spatial negotiation is a sequence of proposals (paths in the case of linear infrastructures) which are evaluated by the different actors, and some counter-proposals are made, until a final consensus or compromise is reached. It is *the contingent approach*. The second is *the structural social approach* (or strategic game approach), where actors consider the object of the negotiation only as a justification and a base for settling prestructured interaction systems, where they try to maximize their global gain in the process. This last is probably the most accurate to reality, but it is much more difficult to implement.

The SPAtial Negotiation Simulator (SPANS) we develop after this model is based on a multi-cognitive-agent approach, in which each modeled actor is implemented in an agent (straight fit). We use all the domain of DAI background knowledge to facilitate the dynamic processing of agents interaction. The approach retained in a first phase is contingent. The core idea is to use for each modeled actor a map showing his representation of the solution. This map is not shared. There is also a shared map which shows the different proposals and their overlapping. So when a new proposal is made by some actor, each of the others evaluate it first according to his value system, which is absolute. It gives a political assessment of the proposal ("this proposal damages the environment in this area, but it is acceptable"). But it is balanced by objective constraints ("Oh, it cannot pass through my backyard"). And finally relational elements are considered ("Well, he is my boss..."). At the end of this evaluation, each tough point is marked and a commitment is put on the counter-argument. The foundation of the proposal system is to "handle open doors", which means that permanently the objective of actors is to keep themselves open choices within space, and to prevent others from being blocked. This is why we expect to reuse the SMAALA system as an evaluator of the possibility of changing a solution, according to each actor. Because "having an open door" implies that in the present situation some changes could be accepted without disturbing values or constraints.

SPANS is still in development and it should use the same geographic interface than the mediation support system.

Conclusion and Perspectives

After an analysis of multi-actor spatial planning, which insisted on the contingency of the exchange of spatial representation at stake, we proposed here an inventory of related functional needs. According to it, we shown that at least two ways could be followed. The first concerns spatial problem solving, and is dedicated to least impact area finding for spatial project, in the context of multi-actor planning. The second is related to negotiation, either for supporting it, or for simulating it.

Because of the structural complexity of these issues, and for a sake of cognitive ergonomy, we proposed to use as an integrative paradigm the Multi-Agent Systems. In that perspective, we shortly recall their principles. And describe the three projects we have in this context. The first, more advanced, deals with complex spatial optimization, with multiple criteria, and spatialized decision system. The second shows an instance of agent based CSCW system dedicated to spatial negotiation support. The last, more prospective, build a coherent framework for simulating spatial negotiations. In all those projects two background principles are kept: consider the social and relational issues as fundamental in planning processes, and build the information processing as an answer to recognized needs. Consequently we refuse the reduction of the complexity of problems. But we also show that Multi-Agent Systems can handle most of such problems, as they are themselves intrinsically complex.

So, as a conclusion, we would like to argue in favor of the development of a fully "Multi-Agent oriented GIS", in which any spatial entity would be implemented in an agent, which, according to the context, could be either active or passive. Such approach would have different advantages:

- Intrinsic parallelism
- Open and adaptive systems
- Self-managed interface for the agents
- Migration of agents through networks
- Correspondence with natural mental representation of human users
- Natural organization for spatial problem solving

According to the present trend toward active object oriented GIS, we presume that our wish will soon be reached. .

Glossary and Abbreviations

Spatial Planning
: Social and political process that aims at deciding where and how in space a project should take place.

Actor
: Person taking part or being interested in a process of spatial planning.

Expert
: Actor, reputed to be proficient in some domain, and who is asked to commit into an assessment of a situation.

(D)AI
: (Distributed) Artificial Intelligence

CSCW
: Computer Supported Collaborative Work, like Groupware

DBMS
: DataBases Management Systems

GIS
: Geographical Information System

KBS
: Knowledge Based System

DSS

Decision Support System

MBMS

ModelBases Management Systems

MCDA

Multi-Criteria Decision Aid

NSS

Negotiation Support System

CAD

Computer Aided Design

M(R)AS

Multi-(Reactive)-Agents Systems

Acknowledgments

SMAALA and SPANS project are supported by a grant (1993-1996) by the Region Rhone-Alpes , with the partnership of CERREP S.A., a consultancy agency specialized in land-use planning and environmental assessment.

Best thanks to Dominique MICHELLAND, Yves DEMAZEAU, Pierre DUMOLLARD and all the MAGMA / LEIBNIZ team.

This document was built using Tamaya-Thot from OPERA project , © INRIA -IMAG 1995.

References

- Simon, H.A., 1977, The new science of management decision, Prentice-Hall, New-Jersey.
- Watzlawick, 1978, La réalité de la réalité, Paris: Seuil, translated from: 1976, How Real is Real? Communication, Disinformation, Confusion. New York: Random House.
- Vincke, P., 1989, L'aide multicritère à la décision, Bruxelles: Editions de l'Université
- Ferrand, N., & Demazeau, Y., 1994, Multi-Agents et Multi-Decisions en Amenagement du Territoire, in Actes des Journées 1994 du Programme Environnement du CNRS, Montpellier: CNRS.
- Ferrand, N., 1995, Multi-Reactive-Agents paradigm for spatial modelling, contribution to the European Science Foundation GISDATA program, Spatial Models and GIS, Stockholm: ESF-GISDATA, to be published: London:Taylor & Francis
- Ferrand, N., & Michelland, D., 1995, Intégration expertise/décision dans l'analyse spatiale des projets d'infrastructures : instrumentation multi-agents, in Proceedings of the 3d. CASSINI program workshop, Marseille: CNRS
- Koning, J.-L., et al., 1995, A multi-agent approach for mediation support on the net, in proceedings of DIMAS95, Krakow:
- Raiffa, H. 1982. The Art and Science of Negotiation, Cambridge: Harvard University Press
- Dupont, C. 1990. La Négotiation - Conduite, Théorie, Applications. Paris: Dalloz.

Nils FERRAND
Team MAGMA

LEIBNIZ LAB.
46 Avenue Felix Viallet
38031 GRENOBLE CEDEX
e-mail to: nils.ferrand@imag.fr

Optimal Field Management for Regional Water Quality Planning

Babak Negahban, C.B. Moss, J.W. Jones, J. Zhang, W.D. Boggess, K.L. Campbell

ABSTRACT:

The process of regional planning can be thought of as an attempt to achieve certain regional goals by implementing local modifications in activities. The development of computer software and hardware has enabled the efficient organization of an extensive amount of local (small scale) information which can be combined into a regional (large scale) plan. There is an exceedingly large number of decision choices to be made at the local level because of the number of local fields (land units), the spatial variability in fields, and the combinations of land use and management practices that are possible. Optimization is a tool which can sift through the numerous combinations of local choices to pick those which, when combined, will produce an optimum plan which best meets regional goals within the constraints imposed on combinations of activities.

The Lake Okeechobee Agriculture Decision Support System (LOADSS) is a GIS-based system which allows planners to assign any of over 100 management practices on any of over 8000 fields in the Lake Okeechobee watershed in South Florida with the goal of controlling excessive nutrient runoff into the lake. Since the number different combinations of management practices on fields are exceedingly large (> 8000^{100}), an optimization module was added to LOADSS to automatically assign practices to fields in order to best meet stated regional goals. Planners can use a mouse and menu-driven user interface to develop optimization formulations incorporating any of 40 environmental, economic, import or export attributes available in the LOADSS database.

In order to test the effectiveness of using optimization algorithms in LOADSS, a pilot basin was selected for optimal regional planning. Eight optimization formulations were selected to be tested on the pilot basin of which six achieved optimal results. Of the six optimal solutions that were obtained for the pilot basin, all but one were able to improve on the LOADSS base plan for the basin in both economic and environmental terms.

LOADSS requires the ARC/INFO 6.1 (GIS) and GAMS (optimization) software and executes on SPARC architecture workstations.

Babak Negahban
Senior GIS Analyst
Breedlove, Dennis & Associates, Inc.
P.O. Bos 720037
Orlando, Florida 32872-0037

Steve Carver, Steve Frysinger and Rene Reitsma

Environmental Modeling and Collaborative Spatial Decision-Making: Some Thoughts and Experiences Arising from the I-17 Meeting

---

This paper describes the outcomes from the initial NCGIA I-17 specialist meeting on Collaborative Spatial Decision-making (CSDM) in relation to the role of environmental modeling and GIS within this field. The meeting, held in Santa Barbara 16-19th September 1995, was attended by 32 participants, many of which had interests or experience in the environmental aspects of decision-making and support. Although much of the meeting was taken up by more general issues, such as developing theories of group decision-making, understanding decision-making processes, methods for multiple representations and design of user interfaces, many issues and problems relating directly to environmental modeling and environmental applications were raised. These included: which models are appropriate for use in CSDM environments; how they can be successfully integrated into GIS and GIS- based Spatial Decision Support Systems (SDSS); the availability of suitable data sets; and problems of multiple representations relating to the compatibility of different model outputs. The general consensus at the close of the meeting was that there was much work to be done in the general arena of CSDM of which identifying and formalising the specific role environmental modeling was a part. This paper presents some thoughts and experiences of three of the I-17 participants in environmentally based CSDM research to help illustrate some of the more pertinent issues and problems involved. Specific issues covered by the paper include: the role of simulation models in environmental decision-making and negotiation setting; the use of GIS and multi-criteria evaluation (MCE) techniques for addressing local environmental decision problems; the role of GIS and MCE in exploratory group decision- making; and examination of laboratory and real world results.

Introduction

Initiative 17 of the NCGIA was recently set up to focus on research issues regarding collaborative spatial decision-making and follows two previous and related NCGIA research initiatives on Use and Value of Geographic Information in Decision-making (I-4) and Spatial Decision Support Systems (I- 6). The setting up of the current initiative by Paul Densham, Marc Armstrong and Karen Kemp had much to do with the realisation that GIS-based Spatial Decision Support Systems (SDSS) form the basis of a potentially powerful set of decision-making tools for problems involving two or more interest groups. It has long been recognised that GIS provides the user with a flexible framework for the development of SDSS (Clarke 1990, Fedra & Reitsma 1990). Work by a number of authors and research teams, including the NCGIA and the Regional Research Laboratory (RRL) Initiative in the UK, has largely identified, and in some cases solved, the main issues and problems regarding GIS-based SDSS for the single user or interest group. The application of such systems in the collaborative decision-making environment brings with it a new set of theoretical, conceptual and methodological problems, thereby opening up a whole new area of research. Particular problems relevant to collaborative spatial decision-making inevitably develop their own distinctive terminology, but include such issues as multiple representations, process intervention, stakeholder empowerment, advocacy, spatial bargaining, group decision-making process, etc. These collaborative aspects of SDSS technology are beginning to appear in a new, but growing literature in the GIS decision-making and support field. The objectives of I-17, as defined by Densham, Armstrong and Kemp, are to: 1. examine the body of theory on the design, implementation and use of computer supported cooperative work (CSCW) environments and evaluate its utility for GIS/GIA; 2. identify impediments to the development of highly interactive, group-based spatial modeling and decision-making environments; 3. develop methods for eliciting, capturing and manipulating knowledge bases that support individual and collective development of alternative solutions to spatial problems; 4. develop methods for supporting collaborative spatial decision-making (CSDM), including methods for managing spatial models; 5. extend capabilities for supporting multi-criteria decision-making in interactive, CSDM environments; and 6. characterise CSDM processes to understand how CSDM technology is and potentially can be used in various CSDM subject domains. To address these issues and problems the NCGIA has set up a working group of 32 people who are actively involved in research allied to this field. A list of I-17 participants and their affiliations is given in Appendix 1. The group, which was drawn mainly from the US but includes representatives from the UK, Switzerland, and Germany, met for the first time in Santa Barbara between September 16-19th 1995. The meeting adopted a flexible structure of small group 'break-out' discussions based around loose topic areas such as tool development, human computer interaction, institutional issues and multiple representations, followed by wider round-table discussions focusing on particular issues identified by the break-out groups. These discussions revealed the presence of two basic interest groups within the I-17 participants: those with an interest in the theoretical and conceptual implications of CSDM; and those whose interests lay more within the technical aspects of enabling the CSDM process. During the meeting it became apparent that a significant proportion of the research issues and problems being raised were being driven from an environmental applications perspective. Questions identified included: which environmental models are appropriate for use in CSDM applications; how can they best be integrated into GIS and GIS-based SDSS; how do restrictions on the availability of suitable environmental data sets affect the utility of these techniques; and how can the problems of multiple representations be solved in relation to the potential incompatibility of model outputs. These and other issues are dealt with below.

Environmental modeling in CSDM

There are three basic questions regarding the role of environmental modeling in CSDM which may be seen as paralleling those affecting the integration of environmental modeling and GIS in general. These are identified above as: 1. model suitability; 2. model integration; and 3. data availability. Although these issues are generic to all environmental modeling applications in GIS, the multi-user aspects of CSDM create certain unique problems of their own, not least of which is the problem of dealing with multiple representations. This is considered to be so significant as to warrant separate discussion in section 3. The three generic questions regarding model suitability, model integration and data availability are discussed briefly in turn below.

Model suitability

Among the wider issues regarding the suitability of existing environmental models for integration into a GIS framework (e.g. spatial and temporal representations, scales of operation, specification of process linkages, compatibility of model structure with the adopted GIS data model, etc.) are a number of distinct issues which relate particularly to the CSDM context. Research into the role of models in group decision-making (e.g. Kraemer 1985) has shown that models helped to: 1. inform participants of problem constraints; 2. inform participants of marginal differences in policy alternatives; 3. involve participants and secure their commitment to outcomes; 4. separate groups and their positions from the policy problem; 5. set agendas by focusing participants on aspects of the problem; and 6. constrain the scope of the conflict. Kraemer's perspective is consistent with an integrative approach to negotiation in which negotiators attempt to reconcile their divergent interests and achieve joint benefit (Carnevale and Isen 1986) rather than try to maximize individual benefits per se. Integrative solutions are desirable because they contribute to long-term stability of relationships and to organizational effectiveness (Pruitt and Carnevale 1982). Walton and McKersie (1965) argue that integrative solutions to a negotiation problem can be supported by the development of a shared definition of the problem and development of shared information on the requirements of others. Modeling of a natural resource's behavior and sharing the results of such modeling among stakeholders or those participating in the actual negotiation process are expected to contribute to this process. As such, the chosen model should adequately reflect and advance the view(s) of the stakeholder(s) in the group decision-making environment. Many environmental models have been developed for non-decision-making purposes with the result that the majority of models that are used in SDSS are, at best second-hand and at worst inappropriate for the task in hand. With second-hand models there is always the risk that the inputs, mechanisms and results do not match their intended purpose and so require considerable restructuring. This is made all the more difficult in the CSDM environment since there are many stakeholders which can, in extreme circumstances, lead to as many model variations as there are stakeholders. Involving the public in CSDM introduces a further complication in regard to model suitability. This is due to the fact that mental models and even cognitive levels can vary widely between stakeholders giving rise to the need for multiple level systems that may use models and/or user interfaces of varying complexity and sophistication (Watson and Wadsworth 1994). On this point, some authors have noted in the GIS literature that confidence in the results of a GIS analysis is closely related to user understanding of the underlying data and models used (Szajna & Scamell 1993, Heywood et al. 1994). Failure to understand the models being used will inevitably lead to a lack in confidence in the results. This is absolutely critical in the context of CSDM where the attainment of a consensus view or universally accepted compromise solution relies heavily on a global understanding of the decision problem and trust in the tools used.

Model integration

Looking at model integration, the problems faced in CSDM are very much the same as those faced in any other GIS application, and as such include decisions about model/GIS linkages, shared data structures, spatial and temporal components, etc. They are however, greatly complicated by the inclusion of multiple stakeholders and the need to arrive at consensus decisions. These isues are reflected in the current concern in CSDM research with the practical and theoretical aspects human computer interaction and cross model comparisons. How the stakeholder interacts with the computer and database in a CSDM environment exerts a significant influence over the success of the decision- making process. As above in regard to user confidence in model results, if the level of complexity and detail in the system and its user interface do not match the mental models of the user, then the effectiveness of the group decision-making process is likely to be impaired. Again, multiple level systems may be the answer to such problems, but there is always the risk of over simplification where complex environmental problems are concerned. In the case of cross model comparisons, in any CSDM application different models may be used by different stakeholders to forward their own cause or ideas in relation to a common theme or problem. The question therefore arises of how to ensure compatibility between models when integrating their outputs to identify areas of commonalty or compromise solutions. This aspect of model integration has far reaching implications in regard to the problems of multiple representations which are discussed in detail below. On a practical level, the time delays involved in CSDM operations resulting from sharing, executing and comparing real models can be a significant impediment to collaborative decision making, especially if the stakeholders are working from geographically separate locations. The tendency is often for one or more of the decision groups to lose patience and elect not to use models in their work, thereby throwing the whole process into disarray. Such practical issues were not discussed in much detail at the I-17 meeting, but it is suggested here that they represent strong determining factors in the success or failure of CSDM.

Data availability

As with any environmental modeling process, lack of suitable data can severely restrict the usefulness of model outputs. This is particularly true in the case of integrating environmental modeling and GIS, in that many environmental models have been developed outside of a GIS framework. As they stand they work well, but problems usually arise when adapting these models to run within or along side a GIS (i.e. difficulties with data formats, file transfers, decisions about tight or loose coupling, etc.). Even when successfully reworked, many environmental models still suffer from the lack of suitable GIS data sets on which to run. Taking a simple lumped hydrological model as a case in point, it is easy enough to adapt the model spatially by placing a fine grid over the catchment and making each cell assume the form of the original lumped model. The outputs in terms of overland flow, through flow and groundwater flow from up slope cells are used as inputs to down slope cells and so on until the accumulated precipitation inputs reach the lowest point in the catchment basin as stream flow. Major problems begin to arise with this model when trying to identify the spatial variations in key model inputs and parameters such as precipitation, infiltration rates, evapo-transpiration, antecedent soil moisture conditions, etc. Any such problems experienced in this context with a single user or decision maker may immediately be increased several fold when working with two or more stakeholders because of differences in model complexities and data requirements. Taking the hydrological example further in the context of a CSDM problem concerning land use planning in a large catchment; the forestry and farming lobby are likely to be interested in soil moisture conditions and so would require a model and data sets which accurately predict this and related variables in order to make informed decisions about yields, planting and harvesting times, etc. Alternatively, water supply agencies are more likely to be interested in runoff and stream discharge when planning supply regimes, bore hole locations, river abstraction points, reservoir storage, etc. and so would require a model and data sets more focused on runoff volumes than the foresters and farmers. Availability of appropriate meta data with which to fashion and temper the use of certain data sets is also a key problem in CSDM. Such meta data may include information about spatial resolution, data sources, spatial variability of error levels with a data set, lineage, etc. In collaborative situations, stakeholders are much more inclined (particularly in the heat of the discussion) to compile data sets which, while they appear compatible, are in fact not because they are based on conflicting modeling assumptions. This is a tricky enough problem when dealing with a single user, but is multiplied further by the number of stakeholders in collaborative decision-making. Meta data is essential in a CSDM environment and must be readily accessible to all stakeholders to allow them to make intelligent decisions regarding data usage.

Multiple representations

Representation, in the context of the current discussion, refers to manner in which the views of interest groups and individuals (i.e. the stakeholders) are presented to the decision group as a whole. Representation may be via mental (psychological, social, cultural and cognitive), visual or computational means; all or some of which may be presented via appropriate models within a GIS framework. The problem of multiple representations in CSDM refers simply to the difficulties and uncertainties of trying to represent the different interests of two or more stakeholders in a common data space. At the level of the discussion presented above regarding generic problems of model suitability, model integration and data availability, multiple representation present several difficulties concerning model choice, cognitive ability, mental models, human computer interaction, compatibility of model outputs and variable data requirements. Perhaps the main area of difficulty in relation to multiple representations is that posed by the need to cope with differences in the compatibility of different model outputs. Although no real consensus of opinion was reached at the I-17 meeting, it appears to the authors that there are perhaps two levels to the multiple representations problem concerning model compatibility: those where the objectives of the different stakeholders differ so much as to require recourse to different models; and those where the objectives of different stakeholders are similar enough to use the same model but perhaps using different data and levels of emphasis. These are discussed below.

Multiple objectives, single model decision problems

One example used to illustrate this and facilitate further discussion at the I-17 meeting was that of river pollution. The two different stakeholders in this example were presented as a chemical company wishing to discharge effluent into the river at an upstream location and a city water management authority at a downstream location affected by the planned discharges. In this case the aim of the consensus building problem is defining a mutually acceptable level of discharge from the chemical plant into the river (i.e. what contaminants and how much?). At first glance, the objectives of the two stakeholders seem entirely different. The chemical company wants the cheapest means of disposing of its process effluents (i.e. river discharge), whilst the city water managers want to minimise health risks and maximise water quality by keeping discharges to a minimum. However, seemingly conflicting objectives or 'positions' can often be abstracted into a single type of 'interest' (Fisher & Ury 1981). In this case, both the city and the plant have an interest in the 'flow' of the river. Note that this abstraction of positions into interests does not imply that there is no conflict. On the contrary, the uses of river flow as transporter of drinking water and of chemical effluent are, at least at certain levels, mutually exclusive. What is important here, however, is that two seemingly different perspectives can be abstracted into a single set of interests. As a consequence, a single model or set of models could be used to represent the environmental resource and its policies for utilization, thereby creating a common platform for discussion and negotiation. Obviously the city and the plant management would use different utility functions to evaluate the acceptability of proposed policies, but at least a common modeling platform can be formulated.

Multiple-objective, multiple model decision problems

Far more complex situations arise when positions cannot be reconciled into one or more common interests. This would happen, for instance, when the downstream city is planning the construction of a reservoir for the storage of drinking water, in between the plant and the city. Although the city still has an interest in the flow of the river in that it replenishes the reservoir and as such does not want the flow to carry the upstream plant's effluent, it now has an additional interest in that pollution will accumulate in the reservoir. Where in the previous example one could conceive of coordinating effluent release schedules with water intake schedules such that both the city and the plant would satisfy their objectives (never mind the cities farther downstream or the aquatic life in the river), now such coordination becomes almost impossible and entirely new policies; e.g. storing the effluent and shipping it periodically, may have to be designed. As a consequence, the same river flow and water quality model will most likely not suffice anymore for representing the salient features of the decision problem. Although this in itself does not prevent meaningful communication between the partners in the negotiation process, it does imply the integration of other, often vary different models. How to develop CSDM environments and models which are flexible enough to accommodate such dynamic integration's of models is a matter for further discussion and research.

Multi-criterion evaluations

Multi-criteria evaluation (MCE) modeling techniques are a good example of how a single model can adequately cope with multiple representations. There are numerous examples in the literature of MCE models being integrated with GIS to solve a number of site search and suitability analysis problems, including regional land use planning in the Netherlands (Janssen & Rietvelt 1990), nuclear waste disposal in the UK (Carver 1991) and industrial location in developing countries (Eastman et al. 1993). In the nuclear waste example the main stakeholders involved are the nuclear industry, the general public and the environmental lobby; each of which can be assumed to have distinctive objectives in mind concerning the search for a suitable site for a nuclear waste disposal facility. The basic objectives of each stakeholder can be summarised as: nuclear industry (minimise costs and maximise safety); general public (maximise distance and minimise health risks); and environmental lobby (minimise environmental impact and minimise health risks). However, with the ultimate and basic objective of all stakeholders being the same (i.e. the identification of a mutually acceptable site) the decision problem can effectively be addressed by all parties using a single MCE siting model with the differences in the stakeholder's specific objectives being expressed through use of different data sets and weighting schemes. The result from such an approach is three different surfaces describing the suitability of different areas of the UK for nuclear waste disposal according to the specific objectives of the stakeholders. Since these have been drawn using the same model, the surfaces are immediately compatible and so can easily be compared to identify commonalties and individual compromise solutions via simple map overlay techniques.

Developing the concepts further

The authors have derived much of the above discussion from direct and shared practical experience in applying GIS-based modeling techniques to both single user and group spatial decision making problems. One of the key problem areas is learning to cope with multiple representations in CSDM. The addition has been made to the single/multiple objective, single/multiple criteria classification of spatial decision problems provided by Eastman et al (1993) of that difficult group of multiple objective, multiple model type decision problems where the compatibility of different environmental models can create such a problem for model comparison and consensus building. Whilst it is hoped that this discussion helps throw some light onto the problems of environmental modeling in CSDM, a number of unresolved issues and new ideas require further attention. These include the role of simulation modeling in negotiation setting and consensus building, exploratory decision support systems and the role of the Internet in creating open systems for mass involvement CSDM.

Simulation models and negotiation setting

One of the main problem areas highlighted above is that of model compatibility in multiple objective, multiple model decision problems. It is suggested here that one possible means of coping with the incompatibility problem when trying to arrive at mutually acceptable solutions is to use an automated simulation modeling approach. Looking back at the river pollution problem, the difficulty of comparing the model results from the different models employed by the two stakeholders involved may be overcome using a modeling environment to simulate incremental changes in the type and volume of effluents discharged by the chemical company. By re-modeling the environmental and economic aspects of the problem from the point of view of both stakeholders, it should be possible to identify the point when the objectives of the chemical company and the city water managers begin to coincide. Obviously some compromise will inevitably be required from both parties if a mutually acceptable solution is to be found, but at least a simulation approach of this kind may provide the necessary information to support an amicable group decision. This approach relies very heavily on mutual trust and understanding between stakeholders regarding each other's models and the automated simulation model provided by the CSDM. Experience would suggest however, that when two parties with different models try to collaborate, they often cannot convince each other to use the other's models and they may not trust anyone else's models either. In this case, perhaps the only way forward is to use all of the stakeholder's models and craft the CSDM in such a way that side-by-side comparisons are possible to facilitate the negotiation. This can be termed simple negotiation setting by provision of on-screen comparisons of model results.

Exploring the decision space

Evidence cited in the decision-making, operations research and IT literature points strongly towards the view that effective group decision making requires full hands-on interaction with the stakeholders. It is highly unlikely that the stakeholders in most CSDM problems will be fully conversant with the models and GIS being used. If CSDM environments are to prove truly interactive, then the models and data used must therefore be accessible to the users via appropriately pitched Graphical User Interfaces (GUI). In an ideal world, such GUIs to CSDM tools would be both simple and structured whilst being sophisticated and flexible all at the same time. Obviously these criteria are to some extent mutually exclusive. However, some experimental systems have been designed which use third party authoring software to design and build GUIs onto the front of environmental models and GIS packages. This provides a product that any reasonably intelligent individual with a basic knowledge of computers could happily use. For example, open architecture systems are widely used to 'shield' the user from the complexities of using environmental models and GIS for decision-making (e.g. Frysinger et al. 1993, Frysinger 1995). Watson and Wadsworth (1994) provide an example of a complex group of three inter- linked models for predicting the effect of land use change on the ecology, hydrology and rural economy of the Tyne Basin in NE England. In this example the decision support GUI shields the user from the complexities of the GRASS based GIS engine and the models themselves, but at three different cognitive levels: the academic with a direct knowledge of problem; the practising professional working in an allied field (e.g. agricultural economics, nature conservation, water supply, etc.); and the educated lay-person where no in-depth knowledge of the problem is assumed. Despite the multi-level approach to the GUI, the system is still based rigidly around the three models and their required input data sets. An example of a much more flexible modeling environment is provided by Heywood and Tomlinson (1995) who have designed a flexible 'meta modeling' interface for use in environmental decision making. This allows much greater flexibility in model choice for a specific decision problem by giving the option to make an informed choice between several models. The GUI is set up such as to allow users to construct their own decision trees using a simple and intuitive drag-and-drop approach to building decision paths. The resulting CSDM environment is, to a certain extent, both structured and flexible at the same time in providing the users with an interface, data sets and models whilst allowing relatively free rein in constructing their individual decision paths in an exploratory fashion. This ability to 'explore' the effects of spatial decisions in the virtual world of the spatial database is a key aspect of GIS technology. There is nothing new about this; GIS practitioners have been carrying out 'what if?' type analyses for as long as there have been GIS practitioners. What is relatively new is the idea that once the technicalities of GIS and environmental modeling are protected from the user in the kind of easy-to-use GUI described above, then true CSDM through public exploration of the effects and implications of spatial decisions normally made behind closed doors in government and company offices becomes a distinct possibility. Coupled with the information explosion and the widespread popularity of the Internet, then true public involvement in CSDM problems where the 'public' is a major stakeholder may become tomorrow's reality. This may ultimately lead toward a new breed of spatial decision support systems focused on the interactive exploration of spatial ideas rather than using GIS in a more traditionalist constructive or hypothesis testing role. The paper by Heywood and Carver (1994) explores this concept under the banner of Idea Generation Systems (IGS) using the example of a family working together to identify an area in which to buy a new house based on a multiple objective, single model MCE approach. Here spatial data relevant to the problem are combined and 'explored', first as individuals (daughter, son, mother, father, grand-mother, etc.) and then as a single family unit in order to create spatial constructions of 'ideas' regarding where might be a nice area to live in and with which everyone agrees. As this is a siting problem, a single MCE model can be applied across the board without any compatibility related problems, but a similar exploratory IGS approach could equally be adopted to multiple objective, multiple model decision problems via the simulation modeling approach described above.

Non-spatial, non-environmental aspects of the decision space

As pointed out by various researchers of how models are used in environmental management (Kraemer 1985, Dutton & Kraemer 1985, Reitsma 1996), the physical and or spatial aspects of environmental decision-making and negotiation often only represent a small portion of the overall complexity of the problem. For instance, in a study of negotiations of the Colorado River Annual Operating Plan (AOP), Reitsma (1996) points out that whereas physical modeling plays an important role as part of the AOP process, the overwhelming majority of issues of negotiation had nothing or little to do with the physical aspects of the river. Instead, they addressed issues such as the legality of operational plans or future precedence that might be established as a consequence of this year's operational plans. Although from an environmental modeling standpoint these aspects may be rather irrelevant, from the point of view of collaborative decision-making these aspects and how to integrate them with environmental models are of the utmost importance. At this point, it is unclear how traditional environmental modeling and issues of strategic, tactical and legal decision-making can be integrated into a single CSDM environment. Although candidate theoretical frameworks such as coordination theory (Crowston & Mallone 1990) exist, integration of these very different aspects of collaborative problems into a consistent modeling framework remains a big challenge.

Mass media decision-making

Heywood and Carver (1994) extrapolate their work on IGS in the form of a hypothetical (if rather utopian) discussion of the potential for restructuring democracy through direct public involvement in policy formulation and decision-making via CSDM tools on the Internet. This may be considered the extreme end of the CSDM vision, whereby everyone with access to the Internet can be involved in providing stakeholder representatives with direct feedback regarding decisions of local, regional, national or global importance. Charges of elitism can easily be levelled at such a suggestion given the current state of development in the Internet. Such a model of democratic decision-making relies heavily on access to the technology and so despite the meteoric increase in Internet connections, there is still a danger of creating an 'information underclass' of people who, for whatever reason, have no connection to this resource. However, if the Internet continues to develop at the current rate, and providing it does not suddenly self destruct, it may be safely assumed that nearly 100% of the population will have direct home or local public access in the none too distant future. As it is at present, the Internet is insufficiently well developed to allow such radical changes in the way decisions are made. Perhaps a more realistic view of the present and future role of the Internet in CSDM is that of a simple information service regarding important spatial and/or environmental decisions roughly in the mould of suggestions for Local Agenda 21 which came out of the Earth Summit held in Rio de Janeiro in June 1992. Local examples of such applications do exist and they increasingly provide a valuable means of providing information and soliciting feedback from interested parties and individuals. Future roles for Internet based tools in wider public information, simulation, decision support, consensus building, negotiation and decision-making are, however, still a possibility given appropriate political will.

Conclusions

This paper has provided an insight into the current status of research into CSDM. This discussion arises out of the past experience of the authors in this field and from discussions at the recent NCGIA I-17 meeting on this subject. The paper identifies some of the more pertinent issues relating to environmental modeling in GIS-based CSDM systems, including the standard issues of model suitability, model integration and data availability, but also the more specialist topics such as multiple representations and model compatibility. In attempting to address those I-17 objectives relating to identifying barriers to developing working CSDM systems, developing new methodologies and identify potential areas of application, a personal view of work ongoing and potential further developments of the concepts and methods of CSDM is provided. However, whatever the views of the authors, it is clear that a significant number of difficult problems and exciting applications areas exist for those involved in CSDM research. It is hoped that I-17 will help produce workable solutions and applications. Clearly, environmental modeling and GIS-based decision support systems are mutually important research areas which hold great potential for the development of powerful CSDM tools. As global population increases so inevitably will the conflict for the earth's finite natural resources. Discoveries of new resources have not kept pace with development and so the onus falls squarely on the sustainable use of existing known resources. With demand rising relative to supply, CSDM tools will be increasing required to help solve conflicts of interest in resource exploitation and development. Geography dictates through the unequal distribution of natural resources that local, regional, national, international and global disputes over these resources and their use will occur. Spatial models and GIS will increasingly become a key technology in supporting decisions about resource management. The need for robust and sophisticated CSDM tools for aiding these important decisions is both urgent and obvious. The final comments in the above section regarding the potential future role of the Internet in CSDM may seem naive in respect to current political systems and their control over the reality of democratic ideals, but they are intended to provoke discussion and thought as to the potential of this fledgling media. In the near future the Internet may provide global access to a massive array of spatially referenced data sets and more importantly the tools to analyse and use them. The potential, at least, for wider public involvement and the true democratisation of the decision making process is just on the horizon.

Acknowledgements

The authors would like to acknowledge the support of the NCGIA in providing financial

assistance to attend the I-17 specialist meeting, Santa Barbara, September 16-19, 1995.

Appendix 1. List of I-17 participants

Marc Armstrong, University of Iowa

David Bennett, Southern Illinois University

Steve Carver, University of Leeds, UK

David Coleman, University of New Brunswick

Helen Couclelis, University of California, Santa Barbara

Joseph Delotto, NYNEX Science & Technology Inc

Paul Densham, University College London, UK

Brenda Faber, CIESIN/TERRA, Fort Collins

Joseph Ferreira, Massachusetts Institute of Technology

Steve Frysinger, James Madison University

Francois Golay, Swiss Fedral Institute of Technology, CH

Thomas Gordon, GMD, KIT-KI, Germany

Jon Gottsegen, University of California, Santa Barbara

Brit Harris, University of Pennsylvania

Lew Hopkins, University of Illinois

Piotr Jankowski, University of Idaho

Douglas Johnston, University of Illinois

Rachel Jones, Loughborough University, UK

Karen Kemp, University of California, Santa Barbara

David Lemberg, University of California, Santa Barbara

Seymour Mandelbaum, University of Pennsylvania

Timothy Nyerges, University of Washington

Alex Pang, University of California, Santa Cruz

Dimitris Papadias, University of Maine

Thomas Pederson, University of Pennsylvania

James Proctor, University of California, Santa Barbara

Rene Reitsma, University of Colorado

Michael Schiffer, Massachusetts Institute of Technology

Susan Suchocki, Claremont Graduate School

Steve Tomlinson, Manchester Metropolitan University, UK

William Wallace, CIESIN/TERRA, Fort Collins

Jeff Wang, US Environmental Protection Agency

References

Carnevale, P.J.D., and Isen, A.M. (1986). The Influence of Positive Affect and Visual Access on the Discovery of Integrative Solutions in Bilateral Negotiation. Organizational Behavior and Human Decision Processes 37: 1-13.

Carver, S.J. (1991) Integrating Multicriteria Evaluation with GIS. International Journal of Geographical Information Systems 5(3): 321-339 .

Clarke, M. (1990) Geographical Information Systems and Model Based Analysis: Towards Effective Decision Support Systems. H.J.Scholten and J.C.H.Stillwell (eds) Geographical Information Systems for Urban and Regional Planning. Kluwer Academic Publishers: 165-175 .

Dutton, W., and Kraemer, K.L. (1985) Modeling as Negotiation: The Political Dynamics of Computer Models in the Policy Process. Norwood, NJ: Ablex.

Eastman, R., Kyem, P., Toledano, J., and Jin, W. (1993) GIS and Decision Making: Explorations in Geographic Information Systems Technology 4. Switzerland: United Nations Institute for Training and Technology.

Fedra, K., and Reitsma, R.F. (1990) Decision Support and Geographical Information Systems. H.J.Scholten and J.C.H.Stillwell (eds) Geographical Information Systems for Urban and Regional Planning. Kluwer Academic Publishers: 177-188.

Fisher, R., and Ury, W. (1981) Getting to Yes; Negotiating Agreement Without Giving In. New York: Houghton Mifflin.

Frysinger, S.P., Copperman, D.A., and Levantino, J.P. (1993) Environmental Decision Support Systems: An Open Architecture Integrating Modeling and GIS. Proceedings of the 2nd International Conference on Integrating GIS and Environmental Modeling. NCGIA:

September 1993.

Frysinger, S.P. (1995) An Open Architecture for Environmental Decision Support. International Journal of Microcomputers in Civil Engineering 10(2): 123-130.

Heywood, D.I., and Carver, S.J (1994) Decision Support or Idea Generation: The Role for GIS in Policy Formulation. Proceedings Symposium für Angewante Geographische Informationsverarbeitung (AGIT'94) Salzburg: 259-266.

Heywood, D.I., Oliver, J., and Tomlinson, S.J. (1994) Building an Exploratory Multi-criteria Modelling Environment for Spatial Decision Support. P.Fisher (ed) Innovations in GIS 2. London: Taylor & Francis.

Janssen, R., and Rietvelt, P. (1990) Multi-criteria Analysis and GIS: An Application to Agricultural Land Use in the Netherlands. H.J.Scholten and J.C.H.Stillwell (eds) Geographical Information Systems for Urban and Regional Planning. Kluwer Academic Publishers.

Kraemer, K.L. (1985) Modeling as Negotiating: The Political Dynamics of Computer Models in Policy Making Advances in Information Processing in Organizations 2: 275-307.

Malone, T.W., and Crowston, K (1990) What is Coordination Theory and How Can it Help Design Cooperative Work Systems. ACM CSCW 90 Proceedings: 375-388.

Pruitt, D.G., and Carnevale, P.J.D. (1982) The Development of Integrative Agreements. V. Derlega and J. Grzelak (eds) Cooperation and Helping Behavior: Theories and Research. New York: Academic Press.

Reitsma, R.F. (1996) Structure and Support of Water Resources Management and Decision Making. Journal of Hydrology (in press).

Reitsma, R.F., Zigurs, I., Lewis, C., Sloane, A.M., and Wilson, E.V. (1996) Experiment with Simulation Models in Water Resources Negotiations. ASCE Journal of Water Resources Planning and Management (in press).

Szajna, B., and Scamell, R. (1993) The Effects of Information System User Expectations on Their Performance and Perceptions. MIS Quarterly: 493-516

Walton, R.E., and McKersie, R.B. (1965) A Behavioral Theory of Labor Negotiations. New York: McGraw-Hill.

Watson, P., and Wadsworth, R. (1994) The Construction of a Spatial Decision Support System for Land Use Planning. Proceedings of the 2nd GIS Research UK Conference. Leicester: 337-348.

Author Information

Stephen J Carver, Dr

School of Geography University of Leeds, LS2 9JT, UK

Tel: +44 113 2333318 Fax: +44 113 2333308

Email: steve@geog.leeds.ac.uk

Steven P Frysinger, Dr

College of Integrated Science and Technology Integrated Science and Technology Program, James Madison University Harrisonburg VA 22807, USA

Tel: +1 540 5682710 Fax: +1 540 5682761

Email: frysinsp@jmu.edu

Rene F Reitsma, Dr

CADSWES University of Colorado Campus Box 421, Boulder CO 80309, USA

Tel: +1 303 4924828 Fax: +1 303 4921347

Email: reitsma@cadswes.colorado.edu

David A. Bennett, Marc P. Armstrong, and Greg A. Wade

Agent Mediated Consensus-Building for Environmental Problems: A Genetic Algorithm Approach

---

ABSTRACT

Environmental problems often result from the distributed and uncoordinated land use management practices of individual decision-makers that, when taken together, cause significant environmental impacts. To develop feasible and politically acceptable solutions to such problems it is often necessary to foster compromise and consensus among a diverse set of special interest groups who possess overlapping sets of objectives; some quantifiable, some not. The union of these sets forms a criteria space that constrains the set of feasible solutions that may be adopted by resource managers. As interest groups work toward the common goal of a mutually acceptable resource management plan they often require assistance in the development, representation, and analysis of this criteria space. Furthermore, new tools are needed that form explicit links between criteria space and the geographic space that is being managed. This paper illustrates how genetic algorithms are used to construct a link between criteria and geographic space and to evolve mutually acceptable solutions to complex environmental problems. Intelligent agents help decision-makers invoke various criteria and learn from the successes and failures of generated solutions in both geographic and criteria space. This knowledge is used to help users assess the fitness of alternative solutions and generate improved solutions to complex environmental problems.

1.0 INTRODUCTION

Research in cumulative impact assessment illustrates how individually insignificant land use decisions can have significant deleterious impacts on the environment (Johnston *et al.*, 1988; Preston and Bedford, 1988). To properly manage environmental resources in privately owned landscapes it is necessary to understand how these individual decisions affect environmental processes across space and through time. The tools used by resource managers to promote environmental objectives in a privately owned landscape depend largely on education, incentive-based policy initiatives (*e.g.* conservation reserve program) and quasi-regulatory compliance programs (*e.g.* commodity programs). Furthermore, private and public concern about the environmental ramification of land management is only one of many competing issues that must be addressed by land managers. To develop feasible and politically acceptable solutions to environmental problems generated by the cumulative impact of multiple decision-makers it is often necessary to foster compromise and consensus among a diverse set of special interest groups who possess overlapping objectives; some quantifiable, some not. Environmental management, therefore, is often a semi-structured problem that requires a collaborative effort among multiple stakeholders.

The difficulties associated with environmental problem solving have parallels in other problem domains. In practice, spatial problem solving tends to be a semi-structured and collaborative effort that crosses managerial and disciplinary bounds. Geographic information systems (GIS) often lack the problem-specific analytical tools needed to adequately explore the solution space of semi-structured problems. Though spatial decision support systems (SDSS) are often designed specifically to address semi-structured problems, they often lack support for collaborative decision making (Armstrong, 1994). Recognizing the limitations of existing geoprocessing technologies, the National Center for Geographical Analysis (NCGIA) established a new research initiative (I-17) designed to address the technological needs of collaborative spatial decision making (CSDM). In September, 1995, the first I-17 specialist meeting was held to identify technological impediments to the development of group-based spatial modeling and decision making environments. Five research topics were identified for discussion at this meeting (NCGIA, 1995):

1. The development of metaplanning capabilities to elicit, capture, and manipulate knowledgebases that support spatial problem solving.
2. The design and implementation of methods to improve decision-makers' interaction with such spatial analysis tools.
3. The provision of mechanisms that enable decision-makers to evaluate alternative solutions to a problem.
4. The identification, selection and incorporation of methods for resolving spatial conflicts in interactive, CSDM environments, including multicriteria decision making.
5. The characterization of CSDM processes, including but not limited to the specification of task models in specific problem domains.

The research presented here is directed toward research topics 2, 3, and 4. In particular, we consider the utility of two technologies, intelligent digital agents and genetic algorithms, as a means of exploring a multicriteria solution space.

2.0 THE CACHE RIVER WATERSHED STUDY

This research is, in part, motivated by a complicated set of resource planning activities that are occurring in the Cache River watershed in southern Illinois. This watershed is largely privately owned and contains an internationally significant cypress/tupelo wetland (a RAMSAR site). There is considerable concern that this unique wetland community is threatened by agricultural land use practices. In this watershed there are three general classes of decision-makers in the watershed who impact land use patterns and, thus, the wetland. These classes are:

1. Farmers who: 1) want to retain full control over their land (private property rights issue); 2) want to maximize farm revenue; and 3) have concern for erosion control.
2. Conservationists (public and private) who want to conserve the ecological vitality of the wetland community.
3. Regional economists who are looking for ways to diversify and bolster the weak economy of this area through: 1) agricultural; 2) industrial; and 3) recreational opportunities.

A multi-disciplinary research team from Southern Illinois University at Carbondale is investigating the impact of alternative resource policy and management scenarios on the economy, hydrology, and ecology of the Cache River watershed. One goal of this research effort is to develop a land use management plan that is generally acceptable to each of these three classes of decision-makers. To evaluate the acceptability of specific management scenarios it is necessary to trace their effect through economic, sociologic, hydrologic, and biologic systems.

3.0 INTELLIGENT AGENTS AND CSDM

Stakeholders in the context of environmental problem solving represent several interests and bring to the negotiation table different types of training, levels of education, experience with computing technologies, and familiarity with the problem that is being addressed. This differential in knowledge can have important interaction effects. For example, a person knowledgeable about the characteristics of soil maps would likely treat a soil layer in a GIS much differently than a person without such knowledge. In spite of their differences, group members must work together to devise solutions to complex problems. If they are unable to overcome the conceptual and technical barriers that inhibit the effective use of alternative technologies then geoprocessing software cannot be used to its fullest capacity.

One way to provide support in heterogeneous decision-making environments is to provide users with intelligent software agents. Two basic forms of software agents have been identified (Shoham, 1993). Personal agents assist users in the execution of routine tasks such as prioritizing electronic mail or maintaining appointment calendars (see for example Kautz *et al.,* 1994). These types of agents learn from repeated interaction with specific users and use this knowledge to guide subsequent behavior. In the context of spatial decision making one can envision a personal agent that properly interprets fuzzy spatial language (*e.g.*, near *vs.* far) or anticipates the type of information that a decision-maker will require to analyze a new alternative. Autonomous agents, on the other hand, are instantiated with a "belief system" (Shoham, 1993) and possess the ability to act on and react to changing stimuli based on this belief system. Deadman and Gimblet (1994) illustrate how autonomous agents can be used to facilitate recreation management. Edmonds *et al.* (1994) implemented five classes of autonomous agents (user, group, floor, conference, and application agents) to facilitate CSDM.

4.0 GENETIC ALGORITHMS

Genetic algorithms are modeled after those processes that drive biological evolution. Alternative solutions in the search space represent individuals in an evolving population. Characteristics that can be used to evaluate the relative success of individual solutions are stored in classifiers. These classifiers are often implemented as bit-strings that document whether a specific solution possesses a given characteristic (Booker *et al*., 1989; Armstrong and Bennett, 1990). For example, an automobile classifier may record whether a vehicle has the characteristic of "four doors". This information can be used to construct a criterion that suggests that a vehicle possessing four doors is more fit as a "family car" than one that does not. The perfect "family car" will, of course, possess a large set of characteristics and it may be that no one vehicle will meet all criteria. So, how is a vehicle designed such that it approximates, as closely as possible, the ideal family car? A genetic algorithm approach would start with the existing population of vehicles and recombine characteristics that best meet the stated criteria. Genetic algorithms, therefore, explore the solution space by adding new alternatives derived from those existing alternatives considered to be most fit. Fitness in this context is proportional to how well a particular solution meets stated criteria.

Three genetic operators are used to combine existing solutions to create new individuals: cross-over, mutation, and inversion. Cross-over is the most powerful of these operators (De Jong, 1990). The cross-over operation generates two new offspring by duplicating two individuals (parents) and swapping "genetic code" beyond some randomly selected cross-over point. To understand the utility of cross-over consider the example discussed below.

4.1 An Example

A decision-maker believes that an ideal family car must possess the following characteristics: four seats, four wheel drive, large cargo area, and enclosed cargo area. These four characteristics can be mapped to a classifier (Figure 1). A fitness value for automobile design can be defined as the total number of desired characteristics possessed by existing or proposed vehicles. Assume that the known universe of automobiles includes only sport cars, four wheel drive trucks, sedans and station wagons, none of which meet all of the desired characteristics. Realizing that the ideal car design has not been invented (no fitness value equal to 4) the genetic algorithm begins to explore the solution space by creating two new car designs by merging the characteristics of a four wheel drive truck and a station wagon (see Figure 1). The generation of new alternatives continues until some terminating criteria is met (*e.g.*, the decision-maker gets a car design that meets stated criteria, a patent is issued, and lots of money is made selling sport utility vehicles).

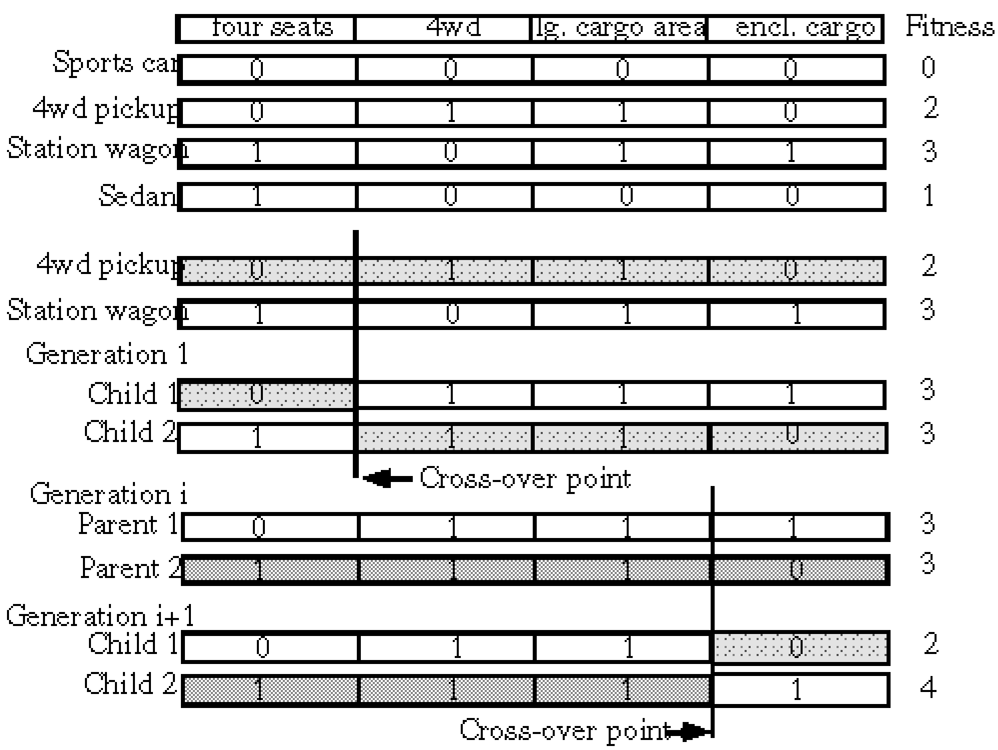


Figure 1. Genetic algorithms evolve new and better solutions through the application of genetic operators (cross-over, mutation, and inversion). The most important of these

operators is cross-over.

4.2 A Formal Description of a Genetic Algorithm

A more formal description of the genetic algorithm is as follows (after De Jong, 1990; Koza, 1994):

1. Generate an initial population, *P(0)*, of solutions. These individual solutions are often created as random combinations of identified characteristics.
2. Do until the termination criteria are satisfied:
   a. For each individual, *m_j*, in the current population, *P(i)*, calculate a fitness, F(*m_j*).
   b. Select *n* individuals from *P(i)* that will be used to generate *n* new solutions for *P(i+1)* via cross-over, inversion, and/or mutation. The probability, *p*, that individual *m* will be used to create new alternatives for population *P(i+1)* is a function of its fitness, F(*m*):

$$p_m = \frac{F(m)}{\Sigma F(m_i)} \qquad (1)$$

   c. Copy the contents of *P(i)* and the newly created alternatives to *P(i+1)*.
   d. Store the "best-so-far" solution.
   e. Remove *m* individuals from *P(i+1)* (based on user defined criteria).
   f. Advance to the next generation *i*+1.
3. Report results to users.

Although geographical applications of this approach are rare, Dibble and Densham (1993) illustrate the utility of genetic algorithms in the solution of location/allocation problems.

5.0 AGENT-DIRECTED GENETIC ALGORITHMS FOR ENVIRONMENTAL PROBLEM SOLVING

As can be seen in the Cache River watershed example, to solve spatial problems it is often necessary to identify management strategies that are acceptable to multiple decision-makers (*e.g.* farmers, conservationists, and regional economists). As suggested above in Section 2, the actions of these decision-makers will be guided by different sets of criteria. The union of these sets forms a criteria space that constrains the set of acceptable solutions. Genetic algorithms are used here to establish a link between criteria space and geographic space and to evolve mutually acceptable solutions to complex environmental problems. Autonomous agents help evaluate how well alternative land use patterns meet user specific criteria and build consensus among competing interests.

5.1 Genetic Algorithms for Two Dimensional Space

Traditional genetic algorithms operate on a finite set of well defined characteristics that are easily mapped to a linear data structure that supports cross-over operations. To use this approach to generate alternative landscapes it is necessary to extend the notion of a linear sequence of genetic code into two dimensional space. Fortunately, the linearization of space is a well studied problem. One of the most useful indexing schemes for accomplishing this task is the Morton sequence which provides a linear indexing scheme for raster-based geographic data sets (Samet, 1990). Using this linear indexing scheme and two randomly selected cross-over points we have a mechanism to create new landscapes that possess characteristics of two parent landscapes (Figures 2 and 3).

5.2 Multicriteria Decision Space

As suggested above, to solve spatial problems it is often necessary to consider the impact of alternative management scenarios on several competing criteria. Furthermore, the set of relevant criteria and the relative importance of specific criteria vary with the goals and objectives of the decision-maker. The classifier and fitness function illustrated in our simple car design example lacks the complexity needed to capture and explore a semi-structured, multicriteria decision space. To overcome these limitations we recast

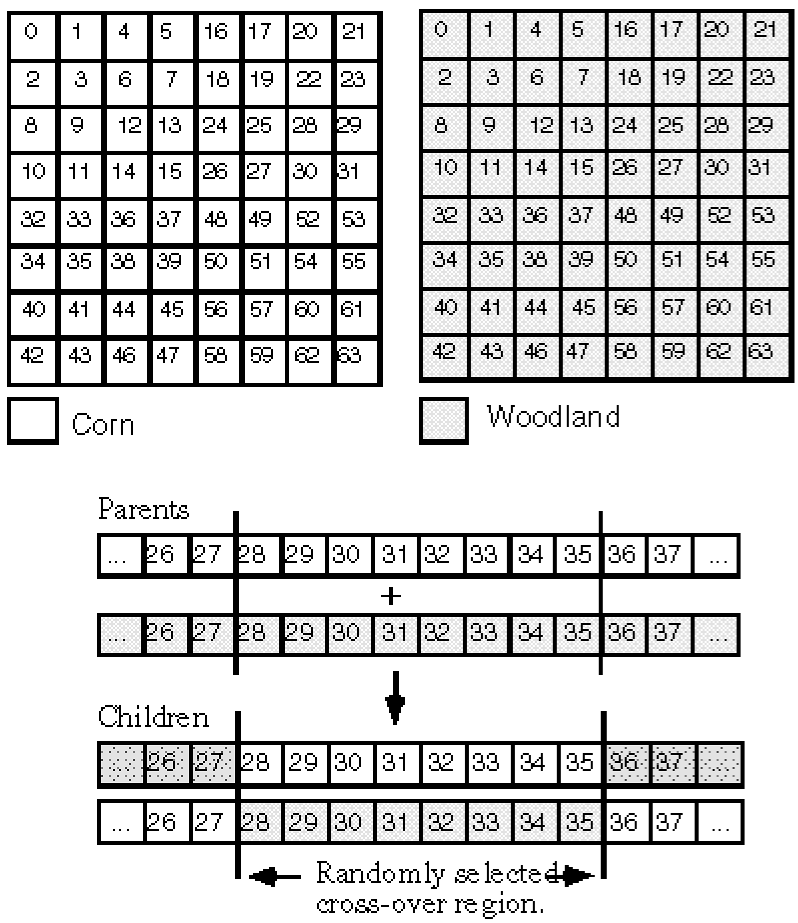


Figure 2. To use genetic algorithms to develop interesting landscape alternatives space is "linearized" using a Morton index.

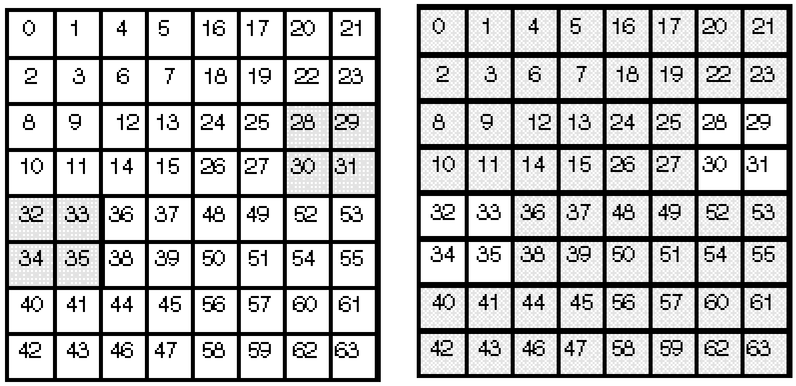


Figure 3. Two new landscapes created by the "genetic cross-over" illustrated in Figure 2.

our fitness function into a modified multicriteria evaluation function. To compare criteria (*e.g.*, minimize erosion and maximize agricultural profit) it is necessary to first standardize criteria scores. This can be accomplished as follows (Carver, 1991):

High score desirable (*e.g.*, economic return)

$$\text{Standardized Score} = \frac{\text{raw score - minimum raw score}}{\text{maximum raw score - minimun raw score}} \quad (2)$$

High score undesirable (*e.g.*, erosion)

$$\text{Standardized Score} = \frac{\text{maximum raw score - raw score}}{\text{maximum raw score - minimun raw score}} \quad (3)$$

In the context of genetic algorithms, the maximum and minimum raw scores must be tracked through the generations. If these values cannot be determined *a priori* then the standardized score must be recalculated for all individuals in the population for each generation. Otherwise, it is necessary to calculate a standardized score only for the individuals added to the population in the current generation. By defining the maximum and minimum scores to be cross generational, and using these scores to standardize multiple criteria, the identification of superior solutions in this genetic algorithmic approach is analogous to an ideal point analysis (see for example Carver, 1991) except that the computer is used to create the set of alternatives that are analyzed. The fitness of landscape *l* is, therefore, calculated as the weighted average of standardized scores.

$$F(l) = \frac{\Sigma \, sc_i(l) w_i}{\Sigma w_i} \quad (4)$$

where:

$F(l)$ = fitness value for landscape *l*.
sc_i = standardized score for criteria *i* given landscape *l*.
w_i = relative weight for criteria *i*.

These criteria weights reflect a qualitative assessment of a single decision-maker or, perhaps, class of decision-makers. One way to calculate a global fitness value for a landscape is, therefore, the mean of independently calculated fitness values:

$$F_g(l) = \frac{\Sigma F_i(l)}{N} \quad (5)$$

where:

$F_g(l)$ = Global fitness value for landscape *l*.
$F_i(l)$ = Fitness value for landscape *l* and decision-maker *i*.
N = Total number of decision-makers.

5.3 Agent Driven Processes

As geoprocessing software becomes more sophisticated, it is able to support the analysis of an increasingly broad set of problems. This richness, however, has a downside: software has become increasingly complex and, thus, more difficult to use. In addition, because of the number of options available, users may not always understand the implications of their choice of a particular method for performing a needed function. In many cases, additional knowledge may be required to support informed use. To reduce the technological burden on the decision-maker, we drive this software system using intelligent agents. At this point, two classes of agents have been implemented, mediating agents and user agents (Figure 4) and the behavior of the agents is straight-forward. A user agent acts on behalf of a specific decision-maker, calculate $F_i(l)$ for each individual landscape *l*, and returns these values to the mediating agent. Using this information the mediating agent calculates $F_g(l)$, selects individuals for cross-over, generates new alternatives and removes unfit individuals from the population.

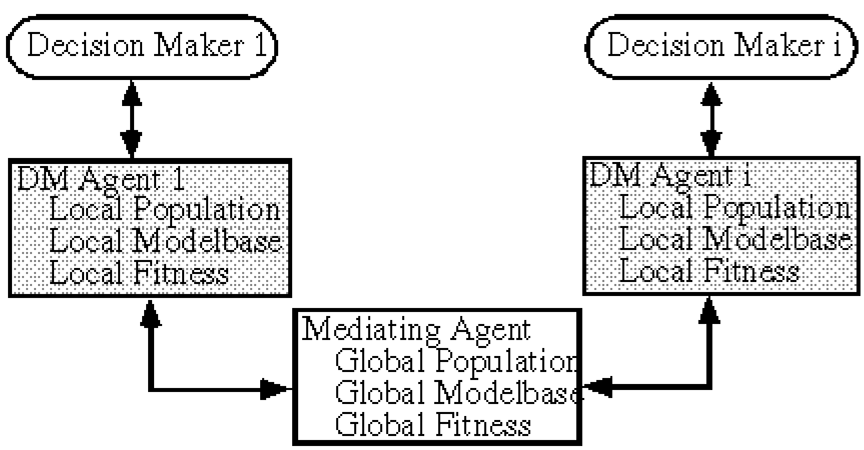


Figure 4. The genetic algorithms are implemented with two classes of autonomous agents: user agents and mediating agents.

6.0 FUTURE DIRECTIONS

With this initial research effort we have implemented a relatively simple and straight-forward integration of genetic algorithms and multicriteria evaluation. From this basic foundation we have identified several avenues for future research. First, we will allow the agents to evolve their own landscapes and contribute locally fit landscapes back into the global population. It is hoped that this modification will accomplish three goals: 1) "allopatric speciation" of landscapes will increase the chances of finding unique and innovation alternatives and create a more diverse "genetic pool" from which to build the solution space; 2) users will have more control over domain-specific models and, thus, more control over the development of alternatives; and 3) by distributing some of the responsibility for the generation and analysis of alternative solutions we will be able to take better advantage of parallel processing technologies. In the next phase of our research we will improve on the agent's ability to learn from and assist users. To address the semi-structured nature of spatial problems the decision-maker must be an integral part of the problem solving process (Densham, 1991). As such, these decision-makers must be able to query the system, develop and apply new models, build management scenarios, engage in "what if" simulations, and purposefully alter the criteria space. These requirements call for intelligent interfaces and modelbase management capabilities. Finally, we want to enhance the ability of the mediating agent to support consensus building. One possibility is to include the development of delta maps in the behavior of mediating agents (Armstrong *et al*., 1992). These maps will focus attention and debate on areas of possible contention and allow the agents to hold constant those areas for which there is agreement.

7.0 CONCLUSION

Research in cumulative impact analysis illustrates how small, individually insignificant disturbances to the landscape can accumulate through space and time to produce significant environmental problems. Yet natural resource management strategies, and the tools that we used to evaluate and implement these strategies, are often directed toward large land areas. To support effective resource management practices new tools are needed that allow us to build consensus among multiple stakeholders and to investigate the cumulative impact of individual decision-makers. This research investigates two technologies that offer promise for such collaborative spatial decision making processes, agent-oriented programming and genetic algorithms. Genetic algorithms are used here to evolve landscapes that meet predetermined criteria. Intelligent agents provide a means of evaluating the fitness of these landscapes based on weighted criteria. Through this interaction between intelligent agents and genetic algorithms management strategies can evolve in such a way that multiple stakeholders are satisfied.

8.0 REFERENCES

Armstrong, M.P. 1994. Requirements for the development of GIS-based group decision support systems. *Journal of the American Society for Information Science*, v. 45, n. 9, pp. 669-677.

Armstrong, M.P. and Bennett, D.A. 1990. A bit-mapped classifier for groundwater quality assessment. *Computers and Geosciences*, 16 (6): 811-832.

Armstrong, M.P., Densham, P.J., Lolonis, P., Rushton, G. 1992. Cartographic displays to support locational decision-making. *Cartography and Geographic Information Systems*, v. 19,

n. 3, pp. 154-164.

Booker, L.B., Goldberg, D.E., Holland, J.H. 1989. Classifier Systems and Genetic Algorithms. *Artificial Intelligence*, v. 40, pp. 235-282.

Carver, S.J. 1991. Integrating multi-criteria evaluation with geographical information systems. *International Journal of Geographical Information Systems*, v. 5, n. 3, pp. 321-339.

Deadman, P. and Gimblet, R.H. 1994. A role for goal-oriented autonomous agents in modeling people-environment interactions in forest recreation. *Mathematical Computer Modelling*, v. 20, n. 8, pp. 121-133.

Densham, P.J. 1991. Spatial decision support systems. In *Geographical Information Systems: Principles and Applications* edited by D.J. Maguire, M.F. Goodchild, and D. Rhind. Wiley, New York, NY.

De Jong, K. 1990. Genetic-algorithm-based learning. In *Machine Learning* edited by Y. Kodratoff and R. Michalski. Morgan Kaufmann, San Mateo, CA.

Dibble, K., Densham P.A. 1993. Generating interesting alternatives in GIS and SDSS using genetic algorithms. In *Proceedings of GIS/LIS '93 ,Volume 1*. Bethesda, MD: American Congress on Surveying and Mapping, pp. 180-189.

Edmonds, E.A., Candy, L., Jones, R., Soufi, B. 1994. Support for collaborative design: Agents and emergence. *Communications of the ACM*, v. 37, n. 7, pp. 41-47.

Koza, J.R. 1994. Introduction to genetic programming. In *Advances in Genetic Programming* edited by K.E. Kinnear. MIT Press, Cambridge, MA.

Kautz, H.A., Selman, B., Coen, M. 1994. Bottom-up design of software agents. *Communications of the ACM*, v. 37, n. 7, pp. 143-147.

Johnston, C.A., Detenbeck, N.E., Bonde, J.P., Niemi, G.J. 1988. Geographic information systems for cumulative impact assessment. *Photogrammetric Engineering and remote Sensing*, v. 54, n. 11, pp. 1609-1615.

NCGIA 1995. I-17: Collaborative Spatial Decision-making. http://www.ncgia.ucsb.edu/research/i17.html.

Preston, E.M., Bedford, B.L. 1988. Evaluating cumulative effects on wetland function: A conceptual overview and generic framework. *Environmental Management*, v. 12, n. 5, pp. 565-583.

Samet, H. 1990. *The Design and Analysis of Spatial Data Structures*. Addison Wesley, Reading, MA.

Shoham, Y. 1993. Agent-oriented programming. *Artificial Intelligence*, v. 60, pp. 51-92.

David A. Bennett, Department of Geography, Southern Illinois University, Carbondale, IL 62901-4514, dbennett@siucvmb.siu.edu.

Marc P. Armstrong, Department of Geography and Program in Applied Mathematical and Computational Sciences, University of Iowa, Iowa City, IA 52242, marc-armstrong@uiowa.edu.

Greg A. Wade, Departments of Geography and Computer Science, Southern Illinois University, Carbondale, IL 62901-4514, wade@cs.siu.edu.

Collaborative GIS in Ecosystem Management System

Jeff Wang
Scientific Visualization Center/Lockheed Martin Service Group
US EPA/National Environmental Supercomputing Center

Abstract

One of the new frontiers of environmental modeling study is ecosystem management decision support. The ecosystem management includes the management of air, water, land use, and demography. In ecosystem management, GIS data play an important role on the decision making.

However, GIS has huge data volume. Even the most advanced centralized storage system can not host all the GIS data for a state or provincial district alone. And no single CPU can process all the numbers in a reasonable time frame, not to mention any analysis or decision support activity. Therefore, GIS data are distributed among different geographical locations, and GIS applications are run on different machines.

Collaborative GIS models are developed to enable data sharing and collaboration via the Internet that connects many computers in the world together. Many issues arise with a model operating in such a vast computer network. Some issues are social in nature, others are technical. This paper intends to provide an in-depth overview of these issues.

This paper will also present a prototype of collaborative GIS in the ecosystem management system, and how the ecosystem management system uses this collaborative GIS model to achieve better environmental management decision support. The paper will discuss the component, architecture, and case study of such a collaborative GIS based ecosystem management system.

Client-Server Approaches to Model Integration within GIS

Raghbir Sandhu & Philip Treleaven

Existing approaches to addressing the problem of integrating modelling methods and Geographical Information Systems range from tightly coupled systems to loosely coupled systems. There is inevitably a trade-off in the linking strategy employed. A balance must be struck between offering high speed data access but remaining GIS-specific on the one hand (e.g. using purpose-built external modules) and lower speed access whilst being able to link to a number of different GIS on the other (using file based interaction, for example).

In this paper, we explore the possibilities presented by employing client-server technologies. At the simplest level, these essentially comprise of a set of operating system supported protocols for handling message passing between applications. Examples include Microsoft's DDE, Apple's Apple Events and sockets for Unix systems.

These have the potential of offering high speed data access yet being able to simultaneously link to different GIS systems. We briefly discuss the techniques themselves and describe a number of new client-server architectures for linking modelling techniques with GIS. An example of such a system is provided by GeoAnalyser, a spatio-temporal analysis tool consisting of a modelling server (which employs Genetic Algorithms for model optimisation) and a GIS client front end.

1. Introduction

GIS are being increasingly used for performing spatial analysis and modelling, moving into territories for which they were not necessarily designed. A variety of statistical methods, models and analytical tools have been used for data analysis, decision support and forecasting in applications such as land-use classification and location selection. This has recently included artificial intelligence techniques such as neural networks, genetic algorithms and expert systems (Sandhu and Treleaven 1995).

This has resulted in a number of *ad hoc* methods of getting around the problem of introducing modelling functionality into GIS systems. Often this has taken the form of separate external packages that are not well integrated with the viewing system. Consequently the linking strategies are wholly incompatible with one another, so that an analysis tool built for one GIS system will not generally accept data output by another.

The spectrum of approaches available for integrating analytical tools within GIS has two extremes (Batty and Xie 1994). In *tightly coupled* systems analytical and modelling utilities exist as specially coded modules within the framework of the GIS. Such modules might be written in a language provided by the GIS (e.g. the Arc Macro Language for Arc/Info). This effectively results in an extension of the GIS's functionality and the analysis system can appear to be seamlessly integrated from a user interface viewpoint. For example, point-and-click queries could be supported in such a system, providing much more intuitive answers for users. Here the GIS would dispatch an *(x,y)* co-ordinate pair to the modelling program, which could then display a result within a window.

Speed of operation is a further advantage of this approach, since all data transfer between the GIS and the modelling subsystem can be performed in main memory.

However, modules built in this way are necessarily *GIS-specific*, since other GIS will generally not support a compatible macro programming language or cater for an identical set of user procedure calls (i.e. hooks). Consequently, there is no hope of being able to run the same models on other GIS. From purely a research perspective, this may not present a problem but in a commercial environment, such compatibility across multiple applications in the same class is highly desirable. For example, the ability to read Lotus 1-2-3 spreadsheets into Microsoft Excel and vice versa is an essential feature for both the users, who inevitably work in a heterogeneous application environment, and application developers, keen to maintain and enlarge market share.

At the other extreme, *loosely coupled* systems interact by using files to exchange data between the GIS and the (separate) analysis package. For example, one might directly read TIGER files into a modelling system and perform some analysis on that data.

The advantage of this approach is that the modelling subsystem can integrate with more than one type of GIS by supporting a number of different data formats. Hence, loosely coupled systems can be *non-GIS-specific*. However, this is achieved at the expense of other desirable features. First, the speed of data transfer between the modelling system and the GIS is much lower, and generally calls for the user to manually export the data of interest before switching to the modelling to read the data.

Secondly, the level of integration is much lower since interaction is at the file level, i.e. it is virtually impossible to implement a common user interface with linked operations. In practice what this means is that point-and-click operations, for example, cannot be supported between the GIS and the modelling program.

Therefore, using current integration methods a trade-off must be made among the three properties discussed above, namely speed of data transfer, GIS-specificity and level of integration. Clearly an alternative strategy is called for which can allow the development of systems with all three of these properties. Such an approach would enable high speed data transfer and permit tight model and GIS integration whilst being portable across a variety of GIS platforms.

A further consideration stems from the observation that end user computer systems are increasingly composed of a set of heterogeneous machines that are connected together via a communications network. For example, a town planning department might have Unix workstations and PCs running on an Ethernet network. Any integration approach adopted should be capable of delivering results despite such a jumble of diverse systems. Obviously, it would be highly desirable if the integration method could in fact utilise these distributed systems for the benefit of the end users.

In this paper, we propose that client-server technologies might be able to satisfy the above set of requirements. This intermediate route has been termed the *co-operative* approach, of which two types have been proposed (Sandhu and Treleaven 1995). First, with *direct co-operation*, the modelling system is directly linked with the GIS via IAC. Secondly, *indirect co-operation* is characterised by the existence of an intermediate interface between the modelling system and the GIS. A workbench would be an example of such an interface. This classification is useful for relating the IAC approach within the spectrum of coupling mechanisms and for distinguishing the type of IAC interaction implemented. We describe and compare the various technologies available and discuss the basic approach used to implement client-server applications.

This work has been motivated by a need to integrate intelligent systems techniques (neural networks, genetic algorithms, fuzzy logic, expert systems, etc.) within the GIS framework, enabling the construction of so-called "Intelligent GIS" (Sandhu and Treleaven 1995).

2. Client-Server Technology

Over the past fifteen years or so, personal and office computing systems have evolved from standalone machines to computers linked by a communications network. This was chiefly driven by the greater flexibility afforded by being able to share valuable system resources, such as file stores, printers, etc. An essential requirement for achieving this shift towards decentralised facilities was a method of passing messages among applications across such a network.

In the software domain, the object oriented methodology has become another enabling force by allowing applications to be built from small, self-contained and reusable components. This has led in turn to far greater flexibility, adaptability and scaleability in the software development process than was possible before. Distributed message passing facilities have also enabled the construction of systems that are object oriented at the *system* level. Examples include client-server databases, which are accessible objects in the system environment.

Utilisation of this technology has not, however, been mirrored in the GIS field. The technology has received very little attention in the GIS literature despite the general shift towards this type of system in the commercial domain. Nevertheless, this state of affairs has not prevented a few GIS vendors from providing limited support for client-server technology in their products (see below).

Several classes of interapplication communication (IAC) technology are available. The technology provides the underlying basis for developing client-server applications that can co-operate across a network.

In the basic client-server model (Berson 1992), a *client* application requests services from a *server* application. An example of such a service might be a request to send some data in an agreed format. The request would be packaged according to the IAC protocol supported and delivered by the operating system. The server application would then verify the request and send back either the data (if the request meets the appropriate criteria) or else generate an error message. The operating system ensures that the messages are delivered successfully and the applications concerned are responsible for dispatching and handling requests for services in the appropriate manner.

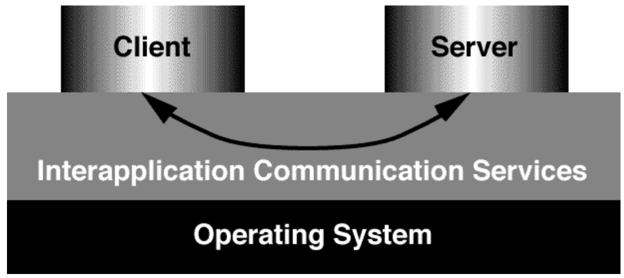


Figure 1 Simple Client Server Model

Client-server systems operate either over a network or on a standalone machine. The advantage of running over a network is that they can make better use of facilities. For example, a fast computer with a large quantity of disk storage is a typical requirement for a database server. Authorised users on the same network as the server can make of use of this hardware, resulting in a more cost-effective set-up.

Here we discuss message passing facilities on three distinct operating systems: Microsoft Windows, Unix and Apple MacOS. These are essentially very similar technologies but they are inherently incompatible with one another and offer different network services.

2.1 Microsoft DDE

The Dynamic Data Exchange (Clark 1992) mechanism of Microsoft Windows implements an operating system-supported protocol for performing message passing between applications. DDE is also available in a similar form on IBM's OS/2 operating system.

A client typically engages in the following procedure to conduct a DDE message exchange (Vose 1990):

- *Allocate shared memory for a DDE object*
  Both applications will access this memory to exchange the DDE message.
- *Create a format for the data to be passed*
  This might be the supplied clipboard format, or any other custom format.
- *Select a type for the information exchange*
  This specifies whether the exchange will be once only, a warm link or a hot link.
- *Send the DDE message*
- *Deallocate shared memory*

The second step is important to note for our discussion. By specifying a special format for the data to be exchanged the structure of the data item is preserved. There is no need to export the data in, say, text format. This helps speed things up considerably provided both applications agree on the format. Note also that the exchange uses main memory - there is no recourse to disk storage - resulting in a high operating speed. The third step permits dynamic data to be transferred - if the data on the server side is constantly changing then setting up a hot link will enable the client using that data to take account of those changes automatically. This gets around the update problem associated with data which have been statically cut and pasted into a document.

While DDE is primarily geared towards seamlessly integrating applications running on the same PC, it can also be used to link applications across a network. Using what are termed *redirectors*, the client initiates a message exchange with a local redirector. This then packages the information and sends it over the network to another redirector on the server machine, which decodes the information back into a DDE message and transfers it to the server application for processing.

Several GIS currently support the use of DDE to various degrees, including Tydac's SPANS MAP, Pafec GIS for Windows and MapInfo.

2.2 Unix Sockets and Pipes

Unix operating systems generally provide at least two methods of communicating amongst applications. Although there are higher level services available from software vendors, our discussion is limited to those available at the operating system level.

A *pipe* is a communication channel initiated by a running process (i.e. application) which behaves very much like standard Unix file redirection. A new process is created in parallel to the existing process and data can be piped to and from it as a sequence of bytes. The restriction here is that data can only be piped to a new process that resides on the same machine as the original process.

The second method, called *sockets*, relies on TCP/IP, the communications protocol underpinning Unix networks and the Internet (Comer and Stevens 1994). A socket is a one-way connection between two processes. Since each socket has associated with it an Internet machine address and port number, the client and server processes can reside on separate machines anywhere on the Internet. Communication is achieved by sending data via a local socket. This is then transmitted over the network to the remote socket from where it may be read by the server process. Typical client-server exchanges involve setting up two sockets at each end, one for sending data and the other for receiving.

Applications written using sockets are fairly portable due to the wide availability of this technology (implementations are even available for the other two operating systems discussed here). However, there are currently no GIS implementations that provide an open interface for integrating using sockets.

2.3 Apple Events

The MacOS provides a number of related facilities for interapplication communication (see Apple Computer 1993 for further details).

At the lowest level, the *Program-to-Program Communications (PPC) Toolbox* provides an efficient mechanism for sending blocks of data between applications that may be spread across a network. This is mainly used by the operating system and applications that need to exchange large quantities of data.

The main IAC facility supported by MacOS is *Apple Events*, which is based on the services of the PPC Toolbox. All events, covering both user events (e.g. mouse clicks and key presses) and system events (e.g. opening an application), are in fact Apple Events. Applications that support Apple Events must be capable of responding to a minimum of four required events and any number of custom events as well as other optional system events.

A preliminary step which must be negotiated by a client application is that of locating the server application. Since the server may also be located on another computer, a standard browser is provided to enable the user to specify which appropriate server application he or she would like to communication with. The process for sending an Apple Event then proceeds as follows:

- The client creates an Apple Event object.
- Data items are packaged into the event object.
- The event object is then sent to the specified server.
- Any data returned from the server are retrieved before disposing of the event object.

An interesting feature is the ability to insert *object specifier records* within an event object. These provide a standard way of referring to particular elements of a document, such as "word 4 on line 2 of paragraph 3 of the document called Letter" or "region labelled London within document UK_populations". If the event is considered to be a verb (e.g. "Copy") the specifier records determine which object that verb should operate on.

Publish and Subscribe uses Apple Events to provide a user-oriented method of sharing dynamic information among a number of applications. From within a drawing package, for example, a user might publish an *edition* of a drawing. Then, from a word-processor, this edition can be subscribed to and the drawing placed within the document. Any consequent changes to the drawing will be automatically reflected inside the text document. This is similar to hot-linking data using DDE but instead relies on files rather than shared memory to perform the exchange.

Custom and system Apple Events can be recorded as a *script*. The script can be executed any number of times to automate simple repetitive tasks. Scripts may also be compiled by hand and this provides a very powerful capability for assembling custom systems that use one or more component applications. The following is an example of such a script for selecting a specific region on a map:

```
tell application "MapMaker"
     select the regions of map MyMap
          whose attribute "population" >= 10000
end tell
```

Scripts can also be used to implement parts of an application's functionality, such as code for responding to a mouse click. It is then possible for the user to replace these scripts with their own, thereby modifying the standard behaviour of an application to suit a particular purpose.

The user is not restricted to using the standard Apple Script language. Through the *Open Scripting Architecture* of MacOS any number of custom scripting languages can be implemented, using language constructs which are appropriate to a particular application domain.

Currently, MapInfo is the only GIS on the Macintosh that supports the use of Apple Events (handled through the Map Basic macro language).

2.4 Summary

The variety and type of IAC facilities on the platforms discussed indicates the lack of standards across the industry. Nevertheless a number of GIS vendors have built client-server interfaces into their systems. This presents the possibility of linking analysis systems to existing GIS using the client-server approach. A number of support services are available in addition to just the message passing protocols; some of these, notably the ability to refer to subelements of a document entity in a standard manner, could be very useful for interacting with spatial data sets accessed via a GIS.

3. Client-Server Architectures for GIS

We have already discussed the basic client-server model, involving an exchange between a single client application and a single server. This simple picture, however, belies the complex structure that distributed systems can exhibit. For example, a client requesting a service from another application could simultaneously act as a server for a third application.

In this section we explore various structures that might be applicable for model integration with GIS.

3.1 Basic Client-Server Architecture

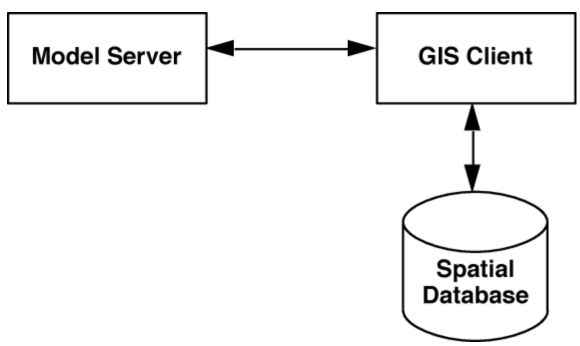


Figure 2 Basic GIS Client-Server Architecture

Using the simple client-server model introduced earlier, one can construct a system similar to that shown in Figure 2, whereby a modelling package is linked using IAC to the GIS. Clearly an essential requirement for a GIS to be used in this way is that it must be IAC-capable. As several off-the-shelf systems have this capability this is not a problem. Data representation is also an issue for which there are two possible solutions. Either the model server must be able to read the raw GIS data or the data must be converted to the appropriate format from within the GIS. The latter can only be adequately handled by a macro language.

The final task is that of integrating the user interfaces and providing execution control to the user. This would involve selecting the appropriate data within the GIS and choosing the type of model to run before performing the analysis. This task would also be carried out through a macro language. At the request of the user, the appropriate macro subroutines are executed, which extract the required data from the spatial database and send them to the model server, where the data are processed and returned to the subroutine for display.

The chief advantage of this arrangement is its simplicity. A certain degree of flexibility is achieved through the use of the built-in macro language and, additionally, it is quite possible for several types of GIS to integrate with the same model server provided the data conversion is done from within the GIS.

We have built an experimental Macintosh-based application for spatio-temporal analysis using the above arrangement. The *GeoAnalyser* system is composed of two components, *GeoClient* and *GeoServer*. GeoClient is responsible for data management and display of spatial information. The display options include thematic maps, data tables (from where the data may be altered manually) and time series graphs. Movies may also be displayed to show the time evolution of spatial variables plotted on a thematic map. Besides stepping forwards and backwards in time over the supplied data, the user is also presented with the facility to project *forwards* in time using a spatial model based on the nonlinear logistic growth equation. This modelling component is contained within GeoServer. GeoClient also contains a Genetic Algorithm to optimise the model parameters for a particular set of data.

Upon initialisation, GeoClient requests the user to identify the server application (which must be running somewhere on the network) and it then dispatches the necessary data to that GeoServer. When a model run is requested, GeoClient sends the appropriate message and model parameters to GeoServer and the newly generated set of spatial data is returned to GeoClient for displaying. Except for the initial stage where the server application must be specified the entire process is completely transparent to the user since the modelling interface is contained wholly within GeoClient.

3.2 Modelling Workbench

A modelling workbench extends some of the capabilities of the architecture presented above. One such structure that might be used to implement a workbench is shown in Figure 3 and discussed in detail below.

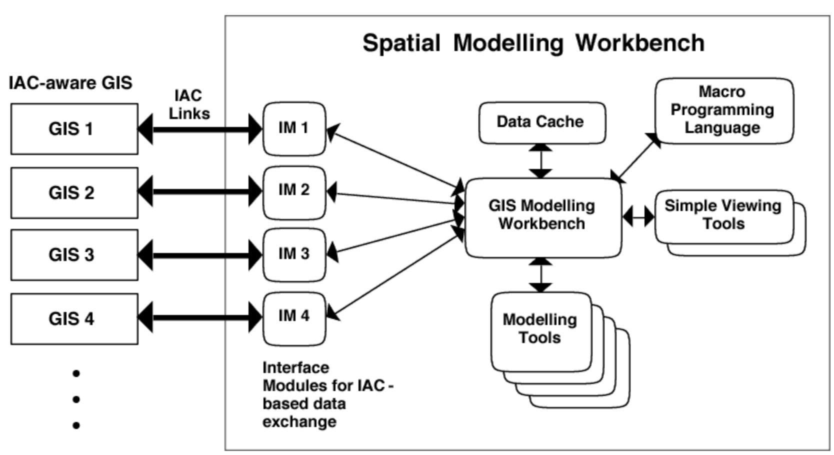


Figure 3 Modelling Workbench

The workbench would be implemented as a single object-oriented application that can co-operate with IAC-capable GIS (these can obviously exist anywhere in the network domain). The workbench has the following sub components:-

- *Interface modules*
  These perform the task of converting spatial data from one format to another and ensure that the workbench is co-operable with several GIS. An API would be provided for creating additional modules to support other formats.

- *Workbench module*
  This is the key subsystem and is responsible for co-ordinating other objects via a unified user interface, executing scripts and organising the sending and receiving of data, using the data cache where appropriate.

- *Modelling tools*
  These objects would perform the analysis on data provided by the workbench module. Links could also be implemented with the viewing tools to display immediately and/or interactively the results of the analysis.

- *Viewing tools*
  Standard view objects such as thematic maps, graphs, etc. would be provided. Custom views could be added through the provision of an API. Linked views could also be supported.

- *Macro programming language*
  Although not strictly necessary, a simple scripting language is very useful for customisation and automation. This would be integrated with the other elements of the workbench.

- *Data cache*
  A data cache is vital to improve overall efficiency by allowing the workbench to reuse data fetched from the GIS in earlier operations.

A system such as this would possess a number of very useful features. First, using interface modules circumvents the need to rely on a macro language built into the GIS for data translation. By providing API's for key functions, additional functionality can be implemented to suit particular tasks. Furthermore, the analysis tools are also well integrated with the display modules within a uniform user interface. The drawback of such a system, however, is that the system is still distinct from the GIS and cannot actually integrate with the functionality and user interface of the host application.

3.3 Data Dispatch Model

The Data Dispatch Model (see Figure 4) attempts to address some of these shortcomings by looking at the design of the GIS as an integral part of the problem. Shown in Figure 4 is an overview of the structure. The system requires no additional application to operate although it would be fairly straightforward to implement a facility for importing live data from a GIS.

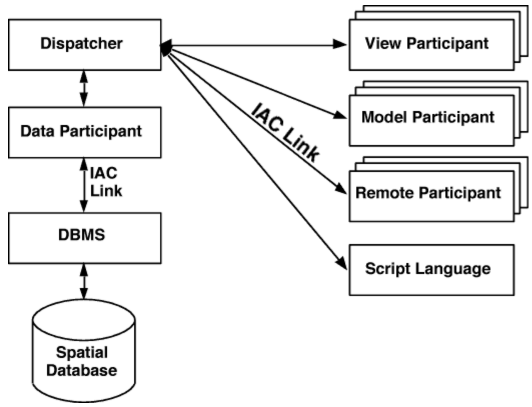


Figure 4 Data Dispatch Model

The system is comprised of the following components:

- *DBMS*
  A database management system is quite common amongst existing GIS for managing spatial data and performing queries on that data.

- *Data Participant*
  This interfaces with the data source and the rest of the system. Being of type *participant*, it can respond to and generate application events (see below). It also contains a cache to help speed up data access.

- *Dispatcher*
  This object is the main control centre of the system. It maintains a list of active participants (which may be spread over a network) and their activity order, i.e. which participants get which particular events. It also permits participants to send messages to each other. For example, if a certain data item is to be altered, then the participant making the change would notify all others to update their state accordingly. An elaboration on this scheme could restrict messages to certain types of participant (for example, to limit a message to all objects of type 'view').

- *View Participant*
  This would implement an object that displays data in some form, which could include maps, graphs, tables, animations, 3D views, etc. Each view would handle standard messages (such as for updating the currently displayed image) and more specialised ones (e.g. to select a certain set of data and show the active selection). Linked views could be supported by issuing the appropriate select and update messages to the other views.

- *Model Participant*
  This would implement the required analysis functions. Model participants have the ability to create new data sets. They can also perform user interaction and issue update messages to views.

- *Remote Participant*
  Any of the above participants can be implemented as a remote service. For example, a model could reside on a separate machine from the rest of the system yet still be fully integrated with the user interface of the Data Dispatcher.

- *Script Language*
  As before, a script language would facilitate the construction of custom systems and automate routine tasks, thereby endowing the system with a greater degree of flexibility.

This system is effectively a GIS in its own right and this is the main criticism of using such an approach. However, by designing the system with the intention of integrating all functions fully with analysis modules, a very effective and extendible system can be developed.

3.4 Summary

Several alternative architectures for integrating modelling and GIS systems have been proposed. An implementation of one of these was discussed. Although the architectures described have their shortcomings, they each solve part of the problem associated with linking modelling tools with GIS. This has been achieved through the deployment of client-server techniques for performing dynamic data transfer between applications.

5. Conclusion

The feasibility of linking modelling tools with GIS depends to a large extent on the type of data transfer facilities provided by the GIS. In the past this was almost exclusively by means of disk files. However recent GIS implementations, particularly those on personal computers, have begun to support the use of client-server techniques as an effective means of transferring data whilst retaining data structure. This has presented the possibility of linking modelling systems with GIS by using these techniques and several possible architectures that might be adopted have been proposed here. However, a greater degree of support for these techniques is required from software houses to achieve full integration.

Whilst the GIS community grapples with this technology, other new technologies are about to change the fundamental way in which we work with applications on computers. For example, compound document technologies (e.g. Apple's recently released OpenDoc) look set to abolish the notion of large multi-purpose applications by encouraging the development of small specialised 'applets.' Such applets might one day blur the boundary that exists between applications, enabling the construction of fully integrated systems that are customised for particular tasks yet remain highly configurable by the user.

Acknowledgements

This research was supported by the Engineering and Physical Sciences Research Council and KPMG Peat Marwick.

References

Apple Computer (1993). *Inside Macintosh: Interapplication Communication*. Addison Wesley.

Batty, M. and Y. Xie (1994). "Modelling inside GIS: Part1. Model structures, exploratory spatial data analysis and aggregation." *International Journal of Geographical Information Systems* **8**(3): pp291-307.

Berson, A. (1992). *Client-server architecture*. McGraw-Hill.

Clark, J. D. (1992). *Windows Programmer's Guide to OLE/DDE*. SAMS/Prentice Hall Computer Publishing.

Comer, D. E. and D. L. Stevens (1994). *Internetworking with TCP/IP - Vol III: Client-server programming and applications*. Prentice-Hall.

Sandhu, R. and P. Treleaven (1995). *Intelligent Geographical Information Systems*. GISRUK'95 (working notes), Newcastle upon Tyne, England, UK. Full paper available via WWW from http://www.cs.ucl.ac.uk/staff/rsandhu/Gisruk.

Vose, M. (1990). Hot Links to Go. *Byte*. November: pp373-377.

Department of Computer Science
University College London
London WC1E 6BT

Tel: +44 (0) 171 419 3694
Fax: +44 (0) 171 387 1397
Email: R.Sandhu@cs.ucl.ac.uk

Applications of 3D Delaunay triangulation algorithms in geoscientific modelling.

Roberto Lattuada* and Jonathan Raper+

*Institute for Animal Health
Ash Road,Pirbright,Woking,Surrey,GU24 ONF
tel. 01483-232441 [352]
fax. 01483-232448
Roberto.Lattuada@bbsrc.ac.uk

+Department of Geography, Birkbeck College
7-15 Gresse Street, London W1P 1PA
tel. 0171-6316470
fax. 0171-6316498
raper@geog.bbk.ac.uk

Abstract

A common method for the reconstruction of a geometric figure given a set of sample points is the use of a triangulation algorithm to connect the points and find the convex hull. In this research, Delaunay triangulation procedures have been used in the reconstruction of 3D geometric figures where the complexity of the problem is much greater than the 2D case. The use of Delaunay triangulations is particularly suited when we do not want to force any constraints on the set of points to be connected. Besides, Delaunay triangulations have some interesting properties as optimal equiangularity and uniqueness (2D).

Introduction and context

During the early 1990's one of the most dynamic fields in spatial information science has been the development of three dimensional GIS [RAP89] [RAP 91]. This early work has generated effective techniques for the visualisation and analysis of three dimensional rasters (based on cuboids or voxels) and for the indexing of vector polylines in three dimensional space [TUR91]. However, by contrast, the interpolation of sparse point datasets into coherent and robust models in unseen, structurally complex geoscientific domains remains a central and largely unsolved problem [RAP 95].

This paper reports research into methods of extending triangulation-based interpolation methods from two to three dimensions, suggests hybrid implementation mechanisms to optimise performance and discusses the particular requirements of geoscientific interpolation as they affect the triangulation procedure.

Algorithms and procedures for 3D Delaunay triangulation

Voronoi diagrams

A systematic approach to the problem of connecting a set of points dates back to 1850 and is due to Dirichlet. He proposed a way to subdivide a given domain into a set of convex polygons. Given two points Pi and Pj in the plane T, the perpendicular to the segment PiPj in the middle point divides the plane T into two regions, Vi and Vj. Region Vi contains all and only the points closest to Pi than to Pj ; if we have more points we can easily extend this concept saying that Vi is the region assigned to Pi so that each point belonging to Vi is closest to Pi than to any other point.

The subdivision of the space determined by a set of distinct points so that each point has associated with it the region of the space nearer to that point than to any other is called **Dirichlet tessellation**.

This process applied to a closed domain generates a set of convex distinct polygons called Voronoi regions which cover the entire domain. This definition can be extended to higher dimension where, for example in three dimensions, the Voronoi regions are convex polyhedrons. If we connect all the pairs of points sharing a border of a Voronoi region we obtain a triangulation of the convex space containing those points. This triangulation is known as **Delaunay triangulation**. An example of the relationship between Voronoi regions and Delaunay triangulation in two dimensions is given in fig. 1. Similarly we can obtain a triangulation for higher dimensions, for example in three dimensions if we connect all pairs of points sharing a common facet in the Voronoi diagram, the result is a set of tetrahedra filling the entire domain.

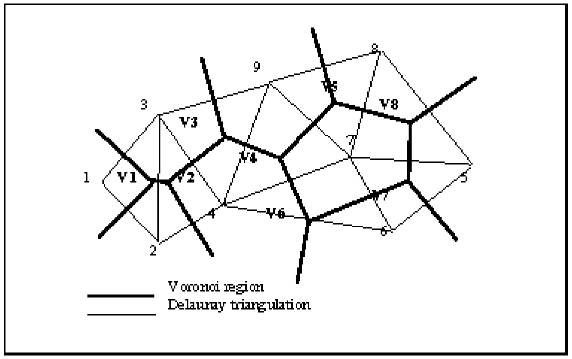


Fig 1 Voronoi regions and associated Delaunay triangulation

The Delaunay triangulation has some interesting properties [BAK87]:

1. IN-CIRCLE
   Given a triangle T(Pi,Pj,Pk) belonging to a Delaunay triangulation DT of a set of points P, no other point of P is internal to the circle defined by Pi,Pj,Pk. As we will see later this property defines a mechanism to automatically build a Delaunay triangulation given a set of points.
2. MAX.-MIN. angle
   Given four points and the associated quadrilateral, the diagonal which splits it into two triangles is optimal in the way that maximises the lesser of the internal angles. This property guarantees that the shape of the triangles is the best possible for that set of points.

Building a Delaunay triangulation

There are a wide variety of algorithms available to build a Delaunay triangulation for a set of points [KG91]. The one described here and used for this research was first described by Bowyer [BOW81] and Watson [WAT81] and has been proved to be the best of those available in terms of quality of elements generated in three dimensions [KAG91]. This is an incremental algorithm, meaning that points are added one at a time into an existing triangulation and, although described for two dimensions, can be easily extended to three or more. The procedure is totally automated and requires no user intervention.

The algorithm can be described as follows:

Let Tn be the Delaunay triangulation of a set n of points , Vn = {Pi | I = 1,..,n}. By defining a simplex as any n-dimensional polygon and a convex hull as the domain to which these points belong to, is formalised as Rs the radius circumscribed to each simplex S of T and as Qs the centre of the n-dimensional circumscribed sphere. Now we insert a new point Pn+1 in the convex hull of Vn and define B={S | S Tn | d(Pn+1,Qs) < Rs} where d(p,Q) is the Euclidean distance between points P and Q. Now B is not empty as Pn+1 lies in the convex hull of Vn and inside a simplex S1 belonging to Tn, so at least S1 belongs to B. The region C formed when B is removed from T is simply connected, contains Pn+1, and Pn+1 is visible from all the points that form the border of C. It is then possible to generate a triangulation of the set of points Vn+1 = Vn {Pn+1} connecting Pn+1 with all the points that form the border of C : this triangulation is a Delaunay triangulation. For a complete demonstration of the above statements see [BAK87].

Degeneracies

It is evident from the definition of a Delaunay triangulation that problems arise in the procedure when certain degeneracies occur in the data.

Common degeneracies in two or more dimensions are :

1. two points are coincident
   The Delaunay triangulation for a set of points is defined only if the points are distinct, so the uniqueness of the points in the given set is treated as a prerequisite for the applicability of this algorithm.
2. three points of a potential triangle are co-linear (or four points co-planar)
   This means that it is not possible to compute a valid centre for the circumcircle of a triangle or the centre of the inscribed sphere for a tetrahedra.
3. four or more points are cyclic

   Common ways to deal with such degeneracy are : reject the point, delay the point insertion, shift its co-ordinates. The choice as to which is the best one often depends on the application; for example rejecting the point may be acceptable if we have a high points density.

For a given set of points in two dimensions, the Delaunay triangulation is univocaly determined and therefore unique, but there are some cases when the triangulation is not unique as there exist different ways of connecting points and all lead to a valid triangulation. This degeneracy is quite common for regular distribution of points, for example in two dimensions when four points lie on a circle and the Voronoi vertexes are coincident.

Data structures

To efficiently handle geometric objects in triangulations there is the need for data structures supporting efficient operations on the objects and capable of handling queries on attributes or, most important, on the object's position within the domain [BAK89b]. The specific data structure implemented in this work is the N-Tree. This is a generic dynamic n-dimensional tree used to recursively divide the domain and provide efficient operations (mainly search) on geometric objects such as points, edges, triangles and tetrahedra.

This kind of data structure also provides an efficient way of checking object uniqueness; for example when an initial set of points is loaded to be triangulated, the points are all inserted in an N-Tree, duplicate points or points that are equal due to the limited machine precision or rounding error will fall into the same space partition, in this way duplicate points are eliminated and the initial requirement for the Delaunay triangulation satisfied.

The N-Tree data structure is similar to the octree data structure, as they are both recursive space subdivision data structures. However in the N-Tree there is no fixed resolution for the subdivision which means that there is no limit to the number of objects present at any moment in the tree. We can see the octree as a way of representing an object where the model and the data structure are completely synonymous since the data structure is in fact the model. In contrast, the N-Tree is more generic (actually a superset of the octree) as its main use is to store objects rather than represent them. This is true for the basic entities such as points, edges, triangles, and so on. Nevertheless, as the Delaunay triangulation of a domain is stored itself in an N-Tree as a set of geometric entities (tetrahedra or triangles), we can also think of it as a way of representing an object, our domain in this case, with the full advantages of a dynamic data structure. This means that the user can regenerate all or part of the model as desired. The scope to generalise and interpolate features by taking full advantage of the N-Tree data structure is currently being investigated.

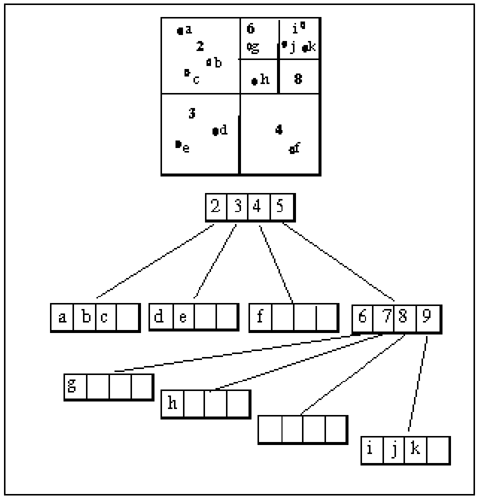


Fig 2 n-tree data structure in two dimensions

Shape refinement

Triangulation of a point set is an important method with many applications including finite element simulations. Though a number of algorithms exist for triangulating a point set in two or three dimensions, few of them address the problem of optimising the shape of the triangular elements [BAR91].

To reduce ill-conditioning in simulations as well as discretization errors, finite elements methods require triangular meshes of bounded aspect ratio. The aspect ratio of triangles or tetrahedra is defined as the ratio of the radii of the inscribing circle to that of the circumscribing one (spheres in case of tetrahedra). While in the two dimensional case the Maximum-Minimum angle property (see above) of the Delaunay triangulation guarantees the best possible shape for the elements, this property does not hold in three dimensions [BAR91]. Though several good heuristics have been published [BAR91] [KAG91], to date there is no known algorithm that triangulates the convex hull of a three dimensional point set with tetrahedra of guaranteed qualities e.g. aspect ratio.

The approach taken in this implementation, based on the Bowyer-Watson algorithm which is sound in two dimensions, is to further subdivide the elements (triangles or tetrahedra) adding points so that the shape of the elements obtained after the split is better than that of the original element.

There are several reasons to try to build well shaped elements, triangulations whose elements have very low aspect ratios are undesirable because :

1. in graphics applications they result in shading irregularities
2. in finite elements analysis they can lead to ill-conditioned numerical problems
3. they weaken the value of the centroid as a representative location.

If we think of our model as based on tetrahedral elements with vertices at the given data points, each point has an attribute which can be associated with the tetrahedra using a local piecewise interpolation method (a simple way is to calculate the average of the values of its vertices). Whichever interpolation method we choose to use, it is important that the shape of the tetrahedra reflect the representative location of the attributes if we want good results from the interpolation process.

Complexity

To define the complexity of the algorithm used we need to highlight which components are dependent on the number of points and therefore play an important part in the definition procedure. When a new point is inserted into an existing triangulation we need to search the list of triangles to find all elements whose circumcircle contains the new point. The time needed to reconnect the points once the elements forming the void are found can be considered constant.

The time needed to triangulate N points can be expressed as : T = (Tk + T1k) where Tk is the time to find the first triangle to be deleted and T1k the time to find all the others.

If we maintain a topological relationship among the elements so that each triangle also stores the reference to all its neighbours we can consider T1k as proportional to the number of elements in the void which is independent from k. With the exception of the first search all other operations are of local nature and may be carried out in a time independent of the number of points currently in the structure. Therefore we can estimate the time complexity of the algorithm as roughly proportional to Tk. The search time Tk will be proportional to k leading to an overall time complexity of O(N2). Using the data structure described in the previous chapter (N-Tree), the cost of the first search can be reduced to O(logN) giving an overall time complexity for the triangulation algorithm of O(NlogN).

Implementation environment

The programming language used in this implementation is C++. This is an object oriented programming language well suited to define both geometric objects and data structures. The advantages of using such approach are :

1. easy extendibility to higher dimensions
2. clean code structure which allows experimenting with different solutions

The modules developed so far are :

1. basic geometric objects both in two and three dimensions (point, edge, facet, triangle, tetrahedra and so on)
2. the N-tree data structure for two and three dimensions
3. Delaunay triangulation in two and three dimensions
4. constrained Delaunay triangulation in two and three dimensions, it is possible to specify a set of constraints (edges, facets) to be included into the triangulation while maintaining all its properties
5. enrichment procedures used to split and improve the average shape of elements
6. triangulation where each sample point has one or more attributes attached to it

Each separate module is arranged as a library that can be easily linked to any application.

Application to 3D geoscientific modelling

Many forms of geoscientific analysis seek to collect data about spatial objects and domains such as features of the solid earth (aquifers), oceans (currents) or atmosphere (weather fronts), which fill or enclose three dimensional space. A complete geometric representation of these domains requires the definition of each known location in a x, y, z co-ordinate system. If a qualitative representation of the object is required then attributes will have to be linked to the geometric descriptions.

The approaches to three dimensional representation and structuring of geo-objects can be categorised as raster, vector, and function based [JON89]. The raster solutions are mostly based around the voxel, which in not necessarily cubic, as a basic unit. Many authors [POI78] [FLO82] have shown the advantages of terrain modelling based on TINs. Compared to the traditional grid based techniques they allow more adaptive modelling and flexible handling of terrain data. The potential of TIN-based methods has been partly exploited in two dimensions but hardly any work has been done in three dimensions.

There are several reasons to try to describe a geo-model using a three dimensional triangulation :

1. the generation algorithm is fully automatic and therefore objective
2. space is uniquely defined and cells are spatially indexed
3. size of elements can be adjusted locally as a function of the complexity of the model
4. the model can be easily edited manually
5. topology is derived from neighbourhood relationships
6. constrained triangulation means we can use vectors or surface constraints (i.e. to represent trends)
7. use of triangular elements is the perfect choice for visualisation since this is the basis for rendering techniques
8. good accuracy and approximation compared to block models
9. integral properties are efficient and easy to calculate
10. we can easily extract from the 3D solid representation of an object the 3D triangulated surface which is its boundary
11. spatial searches and relational queries are easy to implement
12. good performance of Boolean operations

The major challenge of this research is the representation of relationships and interpolation of sparse geoscientific data. Hence, although all the above triangulation algorithms work well with random data sets, when triangulation points are obtained from solid models or boreholes they are not randomly positioned and will usually form highly degenerate set of points aligned on the z axis which form sub-optimum element shapes. This is why a shape refinement post processing has been applied to the initial set of elements (see figure 3a and 3b).

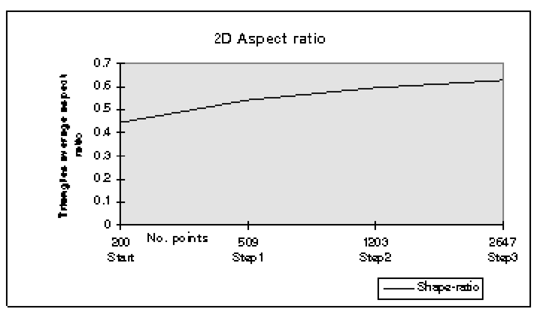


Fig 3a Shape improvement procedure (2D)

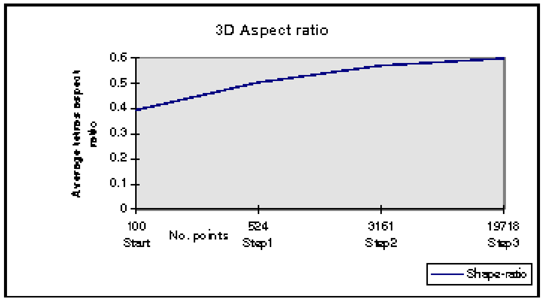


Fig 3b Shape improvement procedure (3D)

Case study

To test the approach set out in this paper a typical geoscientific dataset was triangulated using this method. The data was derived from a set of shallow boreholes drilled in a coastal spit feature. The triangulated points were derived by sampling the recovered sediments at 25 cm vertical intervals and making a laboratory determination of the mean particle size. Data points are arranged vertically forming columns of data points; some data points are missing due to non-sample recovery. Note that the vertical sampling interval is approximately 20-30 times the horizontal borehole spacing. Tests using other methods of interpolation in three dimensions often give poor results when there are such constraints.

The following procedures have been applied to the sample dataset (see figure 4 and figure 5) :

1. shape reconstruction and convex hull definition
2. point set enrichment with linear interpolation for added points
3. linear interpolation of points attributes to associate to each geometric entity (tetrahedra) an attribute value
4. reselection of set of elements based on attribute value

Future development

The procedures set out in this paper seem to satisfactorily show that triangulation methods can be successfully used to structure and interpolate sparse geoscientific datasets. The future aims of this work are to increase the robustness of the implementation and extend its scope to the following areas:

1. The handling of multiple attributes in a full three dimensional geo-relational design;
2. Interpolation between existing samples to define trends and associate values with points added following mesh enrichment;
3. The incorporation of already-created cross sections;

4. The generalisation of shapes represented by tetrahedra.

Further case studies will also be undertaken to optimise these methods for geoscientific research where there is still a relative paucity of procedures for three dimensional interpolation.

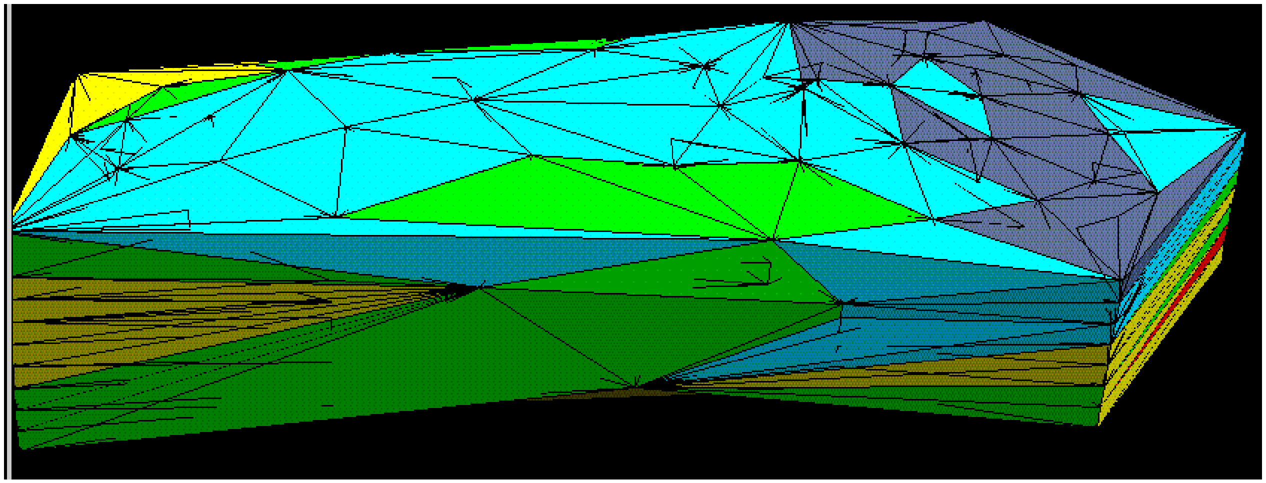
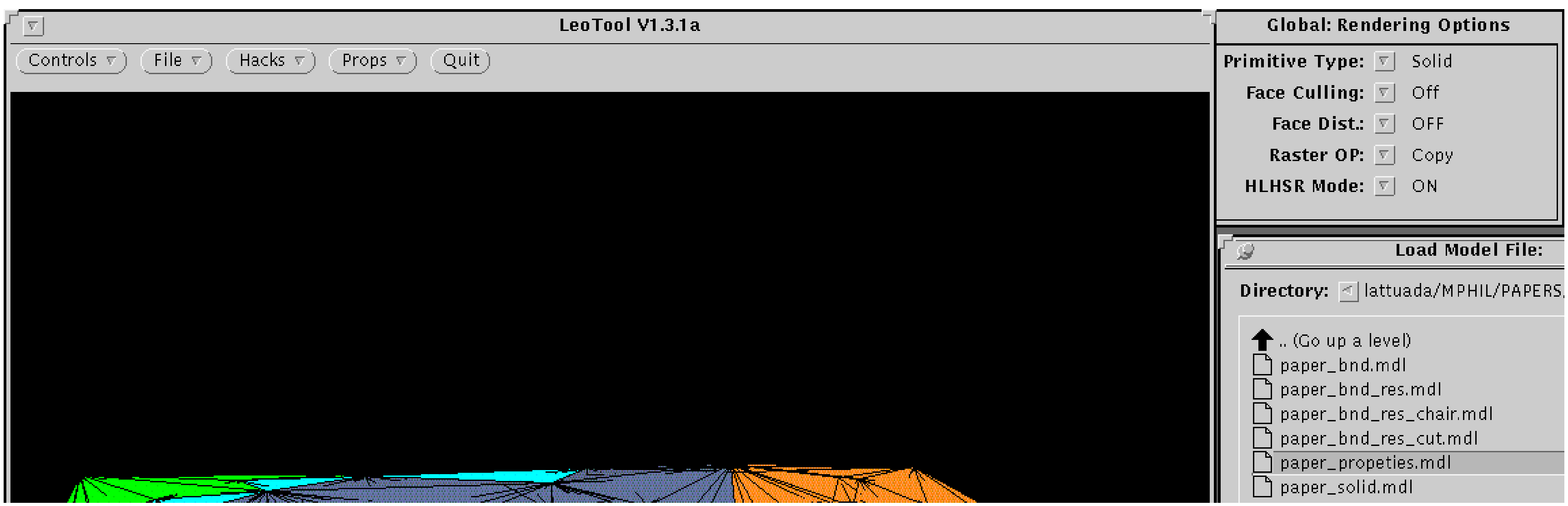


Fig 4 Complete model, attributes are color coded (1).



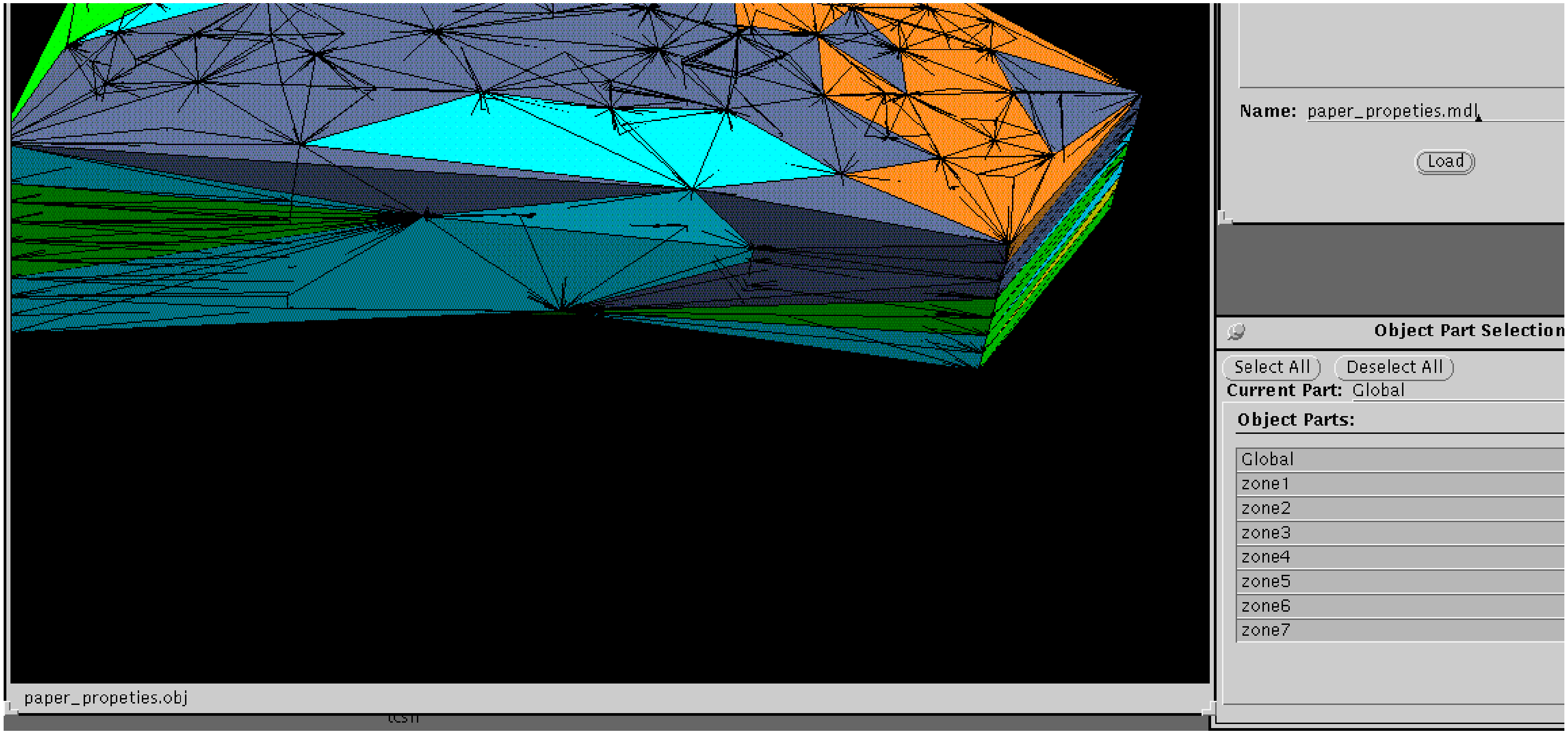


Fig 5 Complete model, attributes are color coded (2).

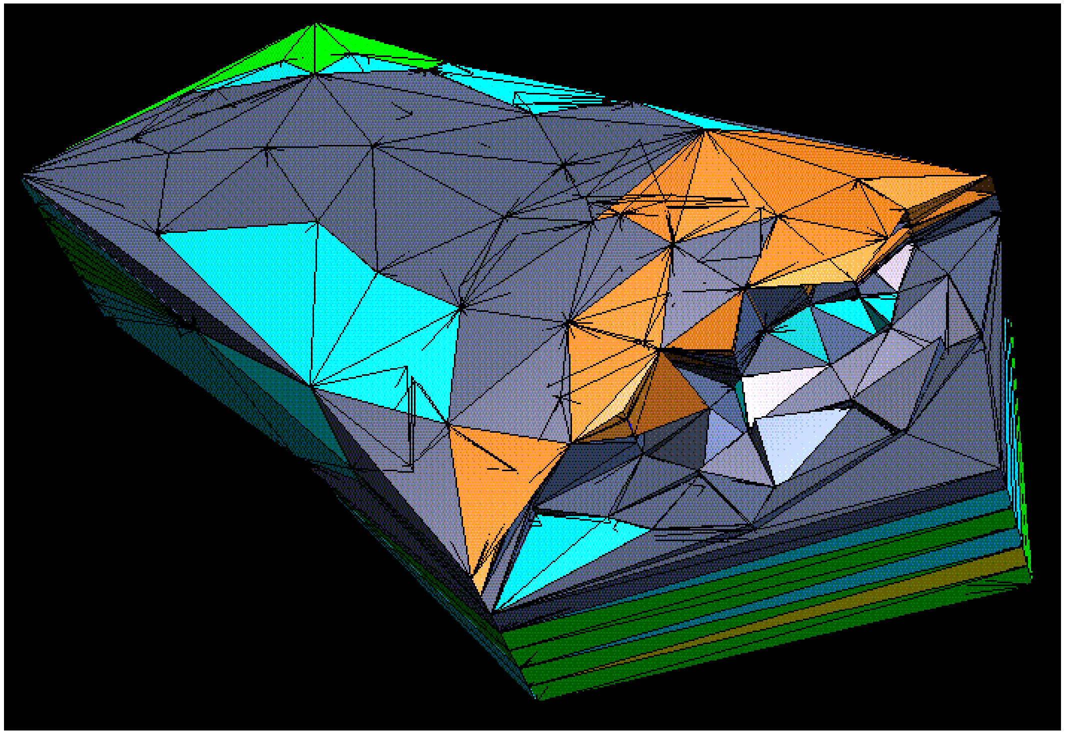


Fig 6 Geometric reselection (chair mode).

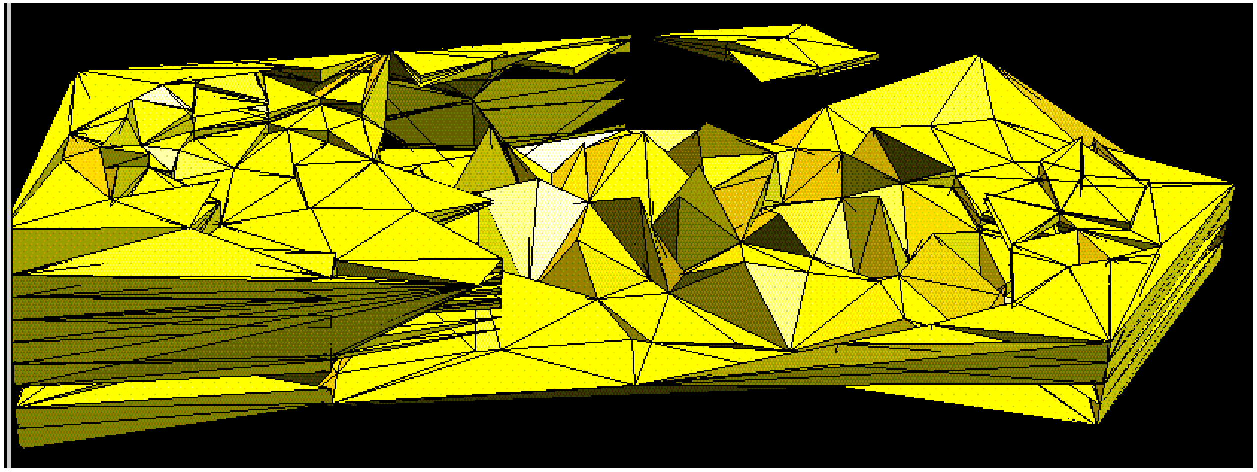
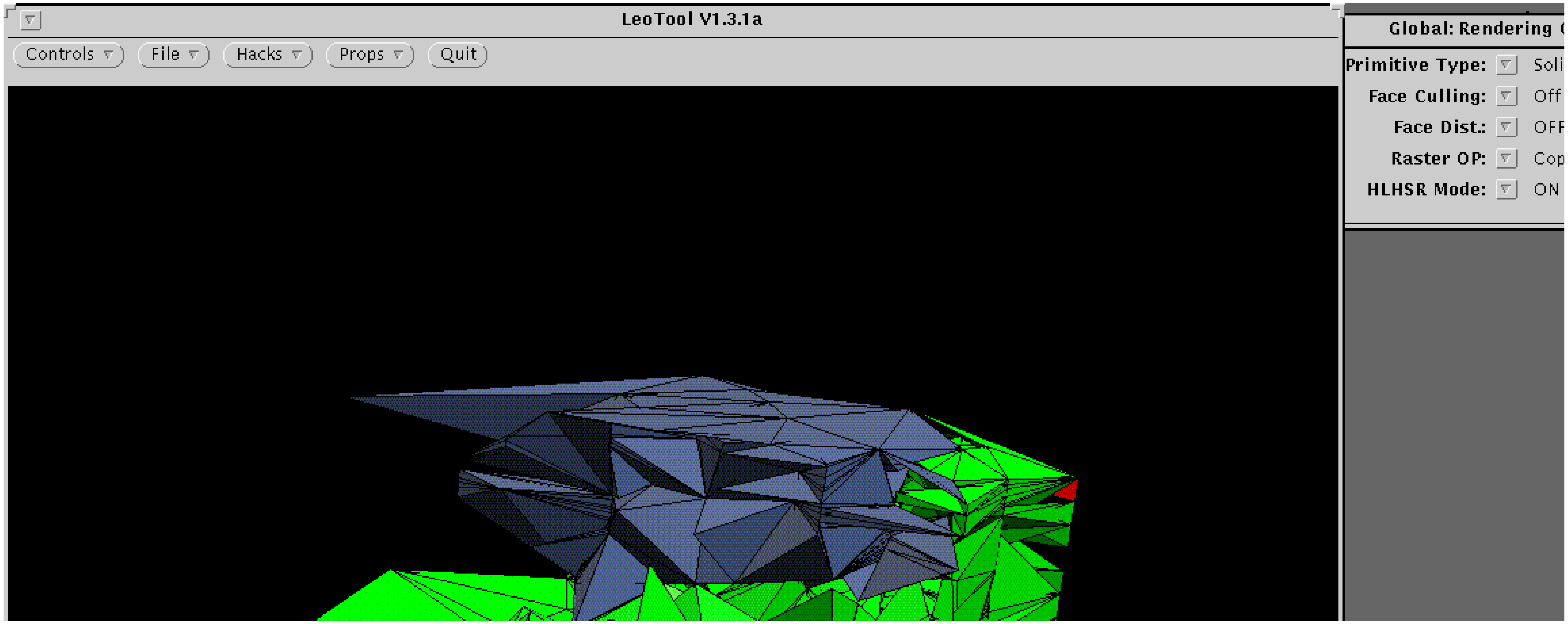


Fig 7 Attribute reselection of 1 layer.



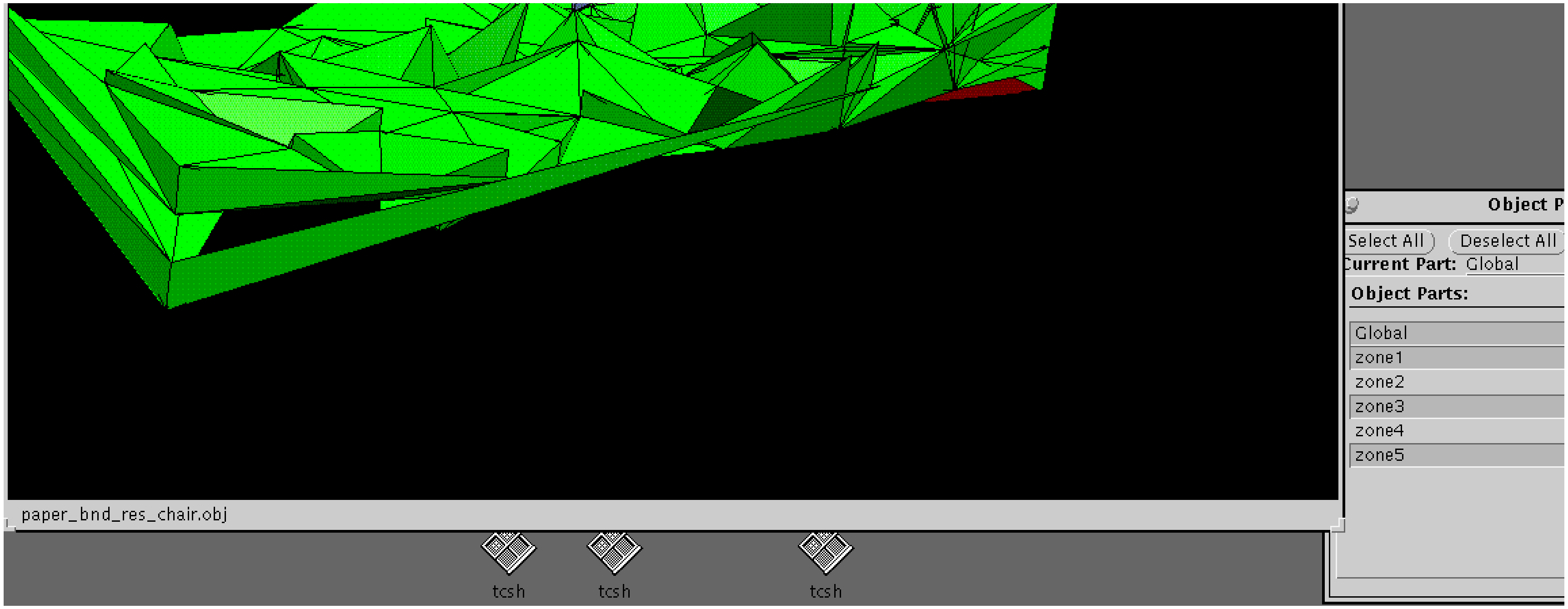


Fig 8 Attribute reselection of 3 layers.

Bibliography

[BAK87] Baker T.J. "Three dimensional mesh generation by triangulation of arbitrary point sets" AIAA 8th Computational Fluid Dynamics Conference 1987

[BAK88] Baker T. J. "Generation of Tetrahedral Meshes Around Complete Aircraft" Second International Conference on Numerical Grid Generation in Computational Fluid Dynamics 1988

[BAK89a] Baker T. J. "Developments and Trends in Three-Dimensional Mesh Generation" Applied Numerical Mathematics (5) 1989

[BAK89b] Baker T. J. "Automatic Mesh Generation for Complex Three-Dimensional Regions Using a Constrained Delaunay Triangulation" Engineering with Computers (5) 1989

[BAK89c] Baker T. J. "Element Quality in Tetrahedral Meshes" 7th International Conference on Finite Element Methods in Flow Problems" 1989

[BAK91] Baker T. J. "Unstructured Mesh Generation and Surface Fidelity for Complex Shapes" AIAA Computational Fluid Dynamics Conference 1991

[BAR91] Barry Joe "Delaunay versus max-min solid angle triangulations for three-dimensional mesh generation" International Journal for Numerical Methods in Engineering, Vol31,987-997, 1991

[BER91] Bern M., Eppstein D., Yao F. "The Expected Extremes in a Delaunay Triangulation" International Journal of Computational Geometry and Applications (1,1) 1991

[BOW81] Bowyer A. "Computing Dirichlet Tessellations" The Computer Journal (24) 1981

[DAV] DavyJ. R., Dew P. M. "A Note on Improving the Performance of Delaunay Triangulation"

[DEY91] Dey T. K., Bajaj C. L., Sugihara K. "On Good Triangulations in Three Dimensions" ACM Communications 1991

[FLO82] De Floriani L., Falcidieno B., Pienovi C. "Triangulated Irregular Networks in Geographical Data Processing", in Rinaldi s. : Enviromental Systems Analysis and Management, pp801,811, 1982

[JAC80] Jackins C. L., Tanimoto S. L. "Oct-Trees and Their Use in Representing Three-Dimensional Objects" Computer Graphics and Image Processing (14) 1980

[JON89] Jones C.B. "Data structures for 3D spatial information systems", International Journal of Geographical Information Systems, 3:15-32

[KAG91] S.Kanaganathan,N.B. Goldstein "Comparison of four point adding algorithms for Delaunay type three dimensional mesh generators", IEEE Transactions on magnetics, Vol 27, No 3, May 1991

[HELLE] Heller Martin "Triangulation algorithms for adaptive terrain modelling"

[LAW86] Lawson L. "Properties of n-dimensional triangulations" Computer Aided Geometric Design (3) 1986

[LEE91] Lee R. C. T., Fu J. J. "Voronoi Diagrams of Moving Points in the Plane" International Journal of Computational Geometry and Applications" 1991

[LOH88] Lohner R. "Some Useful Data Structures for the Generation of Unstructured Grids" Communications in Applied Numerical Methods (4) 1988

[MEA82] Meagher D. "Geometric Modeling Using Octree Encoding" Computer Graphics and Image Processing (19) 1982

[MOO91] Moore D., Warren J. "Bounded Aspect Ratio Triangulation of Smooth Solids"ACM Communications 1991

[PER90] Peraire J., Morgan K., Peiro J. "Unstructured Mesh Methods von Karman Institute for Fluid Dynamics Lecture Series 1990-06

[POI78] Poiker T.K., Fowler J.J., Mark D.M. "The Triangulated Irreguler Network", Proceedings of the Digital Terrain Models Symposium St. Louis, Miss., pp.516-540, 1978

[PRE85] Preparata F. P., Shamos M. I. "Computational Geometry, An Introduction" 1985

[RAP89] Raper,JF (ed.) (1989) Three Dimensional applications in geographical information systems. London: Taylor and Francis.

[RAP91] Raper, JF and Kelk, B (1991) Three dimensional GIS. In Maguire, D Goodchild, M. and Rhind, D. (eds) Geographic Information Systems: Principles and Applications, Harlow: Longman, pp299-317.

[RAP95] Raper, J.F. (1995) Making GIS multidimensional. Proc. 1st Joint European Conf. on Geographic Information, The Hague, Netherlands, 27-31 March 1995, 1, 232-40.

[SAP91] Sapidis N. S., Perucchio R. "Domain Delaunay Tetrahedrization of Arbitrary Shaped Curved Polyhedra Defined in a Solid Modeling System" ACM Communications 1991

[SIB77] Sibson B., Green P. J. "Computing Dirichlet Tessellations in the Plane" The Computer Journal 1977

[SIB78] Sibson B. "Locally equiangular triangulations" The Computer Journal (21) 1978

[TAN83] Tanemura M., Ogawa T., Ogita N. "A New Algorithm for Three-Dimensional Voronoi Tessellation" Journal of Computational Physics (1983)

[TUR] Turner, K. (ed.) Three dimensional modelling with geoscientific information systems , Dordrecht: Kluwer

[VAN91] Vanecek G. Jr. "Brep-Index: a Multidimensional Space Partitioning Tree" International Journal of Computational Geometry and Applications (1,3) 1991

[YER84] Yerry M. A., Shephard M. S. "Automatic Three-Dimensional Mesh Generation by the Modified-Octree Technique" International Journal for Numerical Methods in Engineering (20) 1984

[WAT81] Watson D.F. "Computing the N-Dimensional Delaunay Tessellation with Apllication to Voronoi Polytopes" The Computer Journal (24) 1981

[WEA85] Weatherill P. "The Generation of Unstructured Grids using Dirichlet Tessellations" MAE Rept. No. 1715 Princeton University 1985

[WEA90] Weatherill P. "Numerical Grid Generation" von Karman Institute for Fluid Dynamics Lecture Series 1990-06

About the authors

Roberto Lattuada is currently a second year part-time research student at the geography department, Birkbeck College, London. He is also working as system developer in an EU GIS project for the management of FMD outbreaks. His interests include: GIS applications, mesh generation, software development. He received a degree in Computer Science from the University of Milan (1991).

Jonathan Raper is a lecturer in GIS in the Department of Geography at Birkbeck College, University of London. His current research interests include spatial languages and user environments for GIS, environmental modelling with GIS, GIS and multimedia (he chairs the ESF GISDATA Task Force on multimedia GIS) and 3D/4D modelling in the geosciences (he was editor of '3D applications in GIS').

Hybrid data structures based on Octree and Delaunay Tetrahedral Tessalations (DTT) in 3-D GIS

Quingquan Li and D. Li

Abstract:

In most environmental modelings, geometric information in vertical direction is needed. It has the same significance with horizontal extent, such as in air pollution monitoring. However, most commercial GIS packages are 2-D in these systems; 3-D spatial objects are simplified and represented two dimensionally. For instance, terrain model is represented by TIN. On the other hand, 3-D CAD systems cannot readily be applied to environmental modeling. The development of full 3-D GIS is required in environment projects. Data structure is one of the key problems in 3-D GIS. People have made a lot of results in it, and still have many works to be done. Octree developed from quadtree is thought to be a kind of useful raster structure by most researchers. Delaunay Tetrahedral Tessellations (DTT) developed from TIN is causing more interesting results, in recent years, which may be a powerful vector structure. Each of both structures has different characteristics in spatial manipulation and analysis. In this paper, some introduction about Octree and DTT is presented. Then data compression in Octree and formation of DTT are discussed. A conversion algorithm between Octree and DTT is developed and implemented. Next, a 3-D GIS system based on hybrid data structures is proposed and possible application in environmental modeling is discussed. Finally, the paper concludes with a summary of major findings. 1. Introduction 1.1 Octree and Linear Octree 1.2 Delaunay Tetrahedral Tessellations (DTT) 2. Building and data structures 2.1 3-D run encodinig in Octree 2.2 Formation of DTT based on raster method 3. Conversion between Octree and DTT 3.1 Algorithm 3.2 Implementation 4. A 3-D GIS system based on hybrid data structures 4.1 Framework of system 4.2 Application in environmental modeling 5. Conclusion

Wahun Technical University
Department of Surveying and Mapping
Wuhan, China

CONSTRUCTION AND ROLE OF 3-D GEOLOGICAL FRAMEWORK MODELS

Claudia C. Faunt and A. Keith Turner

Three-dimensional (3-D) digital representations of subsurface geological features materially improve the sophistication and accuracy of environmental process modeling. These geological framework models are especially important when modeling areas of geological complexity. Framework models enhance the modeling process in five ways: 1) by providing input parameters to numerical models, 2) by rapidly visualizing, in 3-D, the results produced by such numerical models, 3) by supporting spatial correlations of these results with geological features and conditions, 4) by incorporating various assumptions concerning the distribution and character of material properties, and 5) by assisting parameter sensitivity assessments using inverse modeling approaches. Given the breadth and potential power of such applications, increased use of 3-D geologic framework models within environmental modeling applications seems assured.

Development of 3-D framework models begins with the assembly of primary data such as: digital elevation models (DEM), geologic maps, geologic cross-sections, and lithologic well logs. Each of these data types can be manipulated by standard Geographic Information Systems (GIS); however the merging of these diverse data types to form a single coherent 3-D digital model requires more specialized software products that have been referred to as Geoscientific Information Systems (GSIS). One method for constructing a 3-D framework model involves four main stages. First, DEM data are combined with geologic maps to provide a series of points defining the outcropping surfaces of individual geological formations. Second, cross-sections and well-logs, properly located in 3-D coordinate space, provide more sparsely defined locations of the same geological units in the subsurface. Third, interpolation of both surface and subsurface information for each geological unit is undertaken with sophisticated gridding and contouring software. Such software incorporates fault discontinuities and produces surfaces defining the boundaries of geological units. However, these surfaces are developed independently, and may extend beyond the proper limits for specific geological units. In the fourth and final stage, a geological framework model is developed when individual surfaces are combined and compared, utilizing appropriate stratigraphic principles to control their intersection and truncation.

No single existing GSIS software product can easily support the entire model construction process when a very large and geologically complex region has to be represented with considerable detail. Several 3-D geological framework models have been developed that represent large regions, at varying scales of detail, within the geologically complex Basin and Range portion of the southwestern United States. These models have used multiple GSIS products during their construction, and involve multiple very large data sets, some exceeding 10-Megabytes, to represent the geological conditions. Successful models have represented areas as large as 100,000-km2 to depths as great as 10-km, involved 20 unique geologic units, incorporated several hundred fault discontinuities, and, in their cellular volummetric representation, over 16-million volummetric cells. Such models appear to approach the

current economic and computer hardware limits, because larger models take significantly longer to develop and require very expensive high-end workstations to visualize or manipulate them.

A prototype of a landscape forest ecosystem management tool using a state transition model and GIS.

Magdiel Ablan, Susan Monteleone, and Miguel Acevedo
Institute of Applied Science and Department of Geography,
University of North Texas, Denton TX 76203.
Email: ablan@unt.edu, susan@gab.unt.edu, acevedo@unt.edu

This paper discusses the development of a forest ecosystem management tool that fully integrates GIS spatial capabilities with the advantages that modeling offers for temporal simulations. The forest model used by this management tool is a landscape-scale state transition model (MOSAIC) which results from scaling-up a gap model (ZELIG) using a semi-Markov framework. Transition states in the model that represent forest successional stages define the GIS cover types.

Individualized scenarios of harvest management practices are developed by the user with a GIS menu-driven interface. Varied menu selections offer the user options to customize the tool to the management application. Menu options provide a mouse- interaction mode for the user to select and define the shape of the harvest area within the forest cover type. Coverages with areas selected are then processed using MOSAIC for a user-defined period of time.

Data analysis includes area calculation for each cover type and covertype diversity indices. Comparison of time steps show forest succession patterns within harvested areas. Differences in cover types between harvested and unharvested scenarios are calculated.

This prototype was developed using data from H.J. Andrews Long Term Ecological Research Forest in the Oregon Cascades.

Britta Allgöwer and Reto Schöning

Forest Fire Modelling with GIS in the Swiss National Park

Introduction

By Swiss federal law (Nationalparkgesetz 1980) the Swiss National Park anticipates the complete protection of nature and its processes. Natural Forest fires belong to the ecosystem of forests. But the Swiss National Park (Engadin valley) is close to human settlements, it is small (170 sqm) and it is - also by law - liable for damages caused by it. In order to fulfil their assignment the park authorities need a fire management policy which allows the individual treatement of each fire situation and which includes ecological as well as economical aspects in the decision.

As forest fires are spatial processes Geographical Information Systems (GIS) are applied as tools to achieve an operational forest fire management system. Since the Division of Spatial Data Handling (University of Zurich) runs the GIS of the Swiss National Park, the people in charge of it work on three topics: (1) implementation of fire spread modelling in Geographical Information Systems, (2) development of fuel models for Switzerland and introduction to the spread modelling and (3) concepts for forest fire management strategies with special emphasis on protected areas.

The overall goal is to provide an interactive Decision Support System which enables the training in a "Fire Simulator" of the people in charge (park authorities, local fire fighters) and the prediction of the damage potential.

Forest Fire Modelling

The basis for the fire behaviour modelling is the Rothermel model for the behaviour of surface fires (Rothermel 1972). It calculates for any given point local intensity and spread parameters for the head of a surface fire. Inputs for the model are a two-dimensional wind field, terrain parameters, fuel moisture and a detailed description of the fuel bed. Based on the local behaviour output by the Rothermel model and on a model for the local shape of fire spread (Anderson 1983), the spread from a set of source locations can be simulated. The influence of barriers (streets, rivers, fuel breaks, etc.) is addressed with a probabilistic model based on the width of the barrier and the flame length. The spread simulation also allows the calculation of the flame length on the entire fire perimeter, which in turn is an important index for the success of various types of fire suppression activities (Rothermel 1983). Once all the required data is available for the Swiss National Park, the model can be used to evaluate different climatic and management scenarios. The fire spread model is implemented in SPARKS, a prototype fire behaviour modelling application. It is fully integrated in a commercial

Geographical Information System (ARC/INFO), built on its raster modelling and applications development functionalities. This allows not only for synoptic analysis over very large areas, but it enlarges greatly the capabilities of the modelling package through the availability of the full range of the GIS' database and spatial analysis functions

Damage Potential

Through the integration of fire behaviour models with GIS models, new insights in the fire danger situation in a management area can be gained. One example is the damage potential that arises from fires starting at a certain point in the landscape. This potential clearly depends on the proximity of the point to sensitive objects and areas like buildings, railway lines, fire-sensitive ecosystems, etc. Proximity is a concept which is used in a great many GIS-related models. However, in the mentioned example proximity can not be modelled as straightline distance, but it must take into account the behaviour of the fire spreading over the landscape. In this approach, the spread simulation is used to calculate the time it takes a fire starting from any point in the landscape to reach an object, under given environmental conditions. This is accomplished by inverting the spread simulation, working from a reached object backwards to all possible sources. The delay times from any point to all objects can then be input to a potential model, used in the GIS realm for assessing accessibility. The model weighs the influence of any reached object on the point's damage potential based on the delay time and the damage susceptibility of the object. The index for the fire damage potential arising from the point is then obtained by simply adding the weighted influences of all objects. This index could be further combined with fire occurrence estimations, probability for early detection, accessibility etc. to give a more complete image of the fire danger situation.

Sensitivity Analysis

Many of the input parameters for the fire behaviour model must be modelled or gathered in extensive field surveys. In order to allocate resources required for the collection of these inputs and to assess the uncertainty introduced in the model results due to uncertain inputs, sensitivity and error analysis with Monte-Carlo Simulation can be performed. This allows the examination - in tabular form or graphically - of the relative importance of each input parameter for a selected output. Also, the uncertainty in the calculated fire behaviour can be calculated for interactively selected points, based on estimated uncertainties of the input parameters.

References

Anderson, H. E. (1983) Predicting Wind-Driven Wild Land Fire Size and Shape. Research Paper INT-305. Ogden, UT: US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, pp. 1-26

Nationalparkgesetz (1980) Bundesgesetz über den Schweizerischen Nationalpark im Kanton Graubünden (Nationalparkgesetz) vom 19. Dezember, pp. 1-3

Rothermel, R. C. (1972) A mathematical model for predicting fire spread in wildland fuels.

Research Paper INT-115. Ogden, UT: US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, pp. 1-40

Rothermel, R. C. (1983) How to predict the spread and intensity of forest and range fires. General Technical Report INT-143. Ogden, UT: US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, pp. 1-53

Authors' address: Britta Allgöwer and Reto Schöning, Division of Spatial Data Handling, Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland, email: britta@gis.geogr.unizh.ch or email: reto@gis.geogr.unizh.ch

Michel ARNAUD, Jean PICHOT

A methodology proposed to obtain a classification of spatial and multivariate data

---

Researchers, engineers' aim is often to class the data they have collected and several variables of wich they have measured. Then they try to form groups, each being as homogenous as possible, and very different from others. They are many algorithms that enable to achieve that aim : Automatic Classification, Dynamic Cluster,...But when observations have a spatial component, that is when data have coordinates in a two dimension space, for exemple, the formely proposed techniques do not take into account this datum. It is yet absolutely necessary to take into account not to lose a very often essential piece of information. In this paper, the authors propose a methodology, that is a succession of tools used one after the other, in complement of the one another. At first, geostatistics, that takes account of the spatial characteristic of data will enable to built a model of variability in a two dimension space and give kriges block estimates. Secondly, the Principal Component Analysis applied to these estimations will take into account their multidimensional characteristic will enable to simplify the space of variables by reducing their number. At last, Automatic Classification applyed to the new factors resulting from PCA will enable may to classify observations objectively. This methodology is easy to include in a GIS and, in the end, we achieve the production of multicritary map easy to be directly read and used for research, management and decision.

George L. Ball,Bernard P. Zeigler,Richard Schlichting,Michael Marefat,D. Phillip Guertin

Problems of Multi-resolution Integration in Dynamic Simulation

Abstract:

Ecological modeling covers many scales of resolution. Each process that we attempt to model must be considered in the context of its own spatial and temporal resolution. If the model involves a single process, the scales can be set accordingly. If multiple models are active in a simulation, the decision of what scale to use can effect either the accuracy of the models or result in severe penalties for computational efficiency. Successfully incorporating multiple scales of resolution, both temporal and spatial, requires some fundamental adjustments in the way modeling is approached. This paper discusses several aspects of handling temporal and spatial dynamics in a multi-scale system, including asynchronous timing; data handling; communication; and visualization.

Introduction

Ecological simulation models have tended to be relatively simplistic. The complexity of the necessary algorithms and the required computing power are a couple of the factors that shape the design and implementation of a model. The modeler has to work, generally, within these constraints and others when deciding what will be included in the final product. The primary approach to modeling is what may be termed data driven design.

Data driven design attempts to construct a model using existing data. The alternative approach, model driven design, has the model being designed with the assumption that the data will be available. Either approach could result in the same answer, but the data driven approach is less flexible for use in other situations. The advantage, however, is to tailor the model to handle the currently available data.

The modeling process becomes even more difficult if we begin to consider multiple interacting models. The tendency to simplify also extends to the choice of examining only one process at a time. Modeling landscape change will require more than a single model unless the major assumption is that no other process has any significant effect (Costanza, this proceeding). There are many examples of this sort of model (e.g. curve numbers for rainfall runoff). Anyone who has had to spend time calibrating one of these models will realize there are other significant factors at work. The goal is to increase the validity of the simulation by incorporating multiple models. As we try to achieve our goal, we find that various tasks in model construction tend to work against us.

In addition to the original data used to drive the simulation, we must also contend with the self-generated data from each model. Each of the various model within a simulation will require storage space for intermediate data as well as output data. With a true dynamic system, the models might also update data in the original "corporate" database. Before we look at possible solutions, we first need to examine what is required to model natural processes.

Ecological Processes

Natural processes all have two characteristics which must be kept in mind when trying to build a model. The first characteristic is time. All processes proceed at a measurable time interval. The second characteristic is space. All processes have a region of influence.

A process by definition extends through time. The total extent is a function of the process itself. We can measure time more precisely than we can measure distance. In either case, we must use an acceptable measure for both if we are to use time and space as variables. An event is something which occurs at a particular time and place. We can define the event by using any three acceptable spatial coordinates (e.g. easting, northing, elevation), and a measure of when (e.g. 1401 on June 21, 1902). How precisely we measure time is related to the instrument used. If we use a cesium clock we can measure time in ridiculously small intervals, but it may not be necessary. In dealing with hierarchical systems, as we move upward in the hierarchy the measure of time increases (e.g. goes from minutes or hours to years or eons). As we move downward through the hierarchy, time speeds up (e.g. going from minutes to seconds). This is more of a relative effect than an absolute one. We could measure everything we observe in say seconds, but then the passage of a year would be a tediously large number (31,536,000 seconds). Measuring the growth of a tree in seconds might require 3,153,600,000 seconds (100 years) before the tree reaches maturity. From a modelling standpoint this would require a lot of calculations which translates to many hours of computer cycles. It is more useful to adopt a measurement of time that is consistent with the process in which we are interested. For example, the spread of a fire across the landscape may take a couple of hours, where as the life cycle of a tree may require several hundred years.

Spatial resolution

Having established a suitable time scale, we need to determine a scale for spatial resolution. Our ability to gather data at a scale of resolution that might be universally useful for all parts of the model would prove very difficult. If we are modelling vegetation processes at the Grand Canyon and measure our vegetation plots in centimeters, do we want to use centimeters in our calculations of elevation? If we use AVHRR data to describe the vegetation, the resolution will be 1 kilometer. At that resolution we can't tell much about the vegetation and the mountains will look somewhat like bumps on the landscape. If we look at the fire and tree example, we find that the fire might cover an area from less than a hectare to several square kilometers. The tree on the other hand, might have a radius of influence measured in meters. If we model the events at a spatial resolution of 1 meter to handle the tree, we see that for the small fire that might be possible. For the larger fire we would potentially need a great deal of storage space to accommodate that much data. If we worked with 1 hectare resolution, we might not even see the tree!

All questions involving the real system must be set in the context of location or time. For instance, if our model has determined that there is a tree present, this has a very different meaning if we are modelling a forest or an arid desert. In the same manner, indicating that a ground water model predicted water transport of 3 cm per year has a different meaning depending on the soil conditions. Time is relative and therefore our models must be consistent with the time scales of the process. Measuring the movement of groundwater over grid cells of 1000 meter resolution and a time step of seconds would not be practical nor particularly revealing.

Related aspects of modeling

Continuous versus discrete modelling refers to how time is represented. We relate events as to their position in time and we can measure this position to some arbitrary level of precision. Time is continuous, meaning that it is marked by an uninterrupted extension in sequence or to put it another way, it has no distinction of content except by reference to something else such as numbers (e.g. seconds, minutes, and hours). Discrete time is a sequence of distinct intervals in which any intervening information is disregarded or assumed to be inconsequential.

Our use of computers requires that we make some decisions about how we represent time. Analog computers operate with numbers represented by directly measurable quantities such as electrical voltage. As the voltage increases or decreases, the computational values change. Electrical voltage can be continuous in nature and therefore some models that can be implemented on analog computers can use continuous time. The digital computer represents numbers directly as digits (ones and zeros) and therefore time is automatically represented as an interval or step. Most modelling is done on digital computers and therefore the models are discrete. mathematical representation using differential equations is an attempt to circumvent this problem. Discrete time steps tend to be the choice of most modelers. It is easy to grasp conceptually and easy to implement. However, since time is referenced to something such as an event, we might measure time advance bo other means.

One approach is to treat time in reference to events. In most models, there is a period of time, however finite, in which the solution is essentially static. Nothing is really happening, therefore we may be doing useless calculations. If we look at only the time period in which something happens we can basically skip over the intervening time period. This approach is known as discrete event simulation (Zielger, 1976). With discrete event simulation, the model only performs calculations when it is ready to change states. There is an inherent synchronization in this approach, since each model will automatically be staged according to the next event time.

We have discussed the idea of multi-resolution time scales in natural processes, now let's look at the spatial resolution. There are two types of model that can be used in natural process simulation: non-spatial and spatial models. Non-spatial models (e.g. carbon cycling in plants) are very prevalent in the literature but are not of interest to this conference.

Spatial models are those models in which two or three dimensions are represented. Some models such as those used in the artificial life simulations, use a synthetic world. The world is usually represented by a regularly spaced grid. The object in the simulation then move from one grid space to another during the simulation. The use of abstract space is fine for Alife but for ecological simulation we need a better representation of the real world.

The use of GIS databases allows us to represent the earth in a manner known as georeferencing. A georeferenced database uses some coordinate system that can be related the surface of the earth. Therefore we can build very detailed models of the earth's surface and use this to drive our models. The amount of detail that the database contains is dependent on the spatial resolution used to gather the data.

In some, if not most cases, the resolution is dictated by the original form of the data (e.g. USGS 7.5 minute Quad). When the data is captured and put into the GIS system it is usually in vector form. This is a very efficient method of storing data about points, lines and polygons. When we use it in simulation models, the data is generally rasterized to a useful resolution. The transformation from vector to raster leads to a significant increase in data storage size. For this reason, simulations using GIS data are primarily chosen to make use of only small parts of the original data if high resolution is required, or lower resolution of the entire database if large scale (landscape level) simulation is desired.

As was stated earlier, each model has its own unique spatial and temporal resolution. In a very complex landscape dynamics simulation, the data may be a mixture of high resolution and low resolution driving various models. As each model progresses, its data are altered and perhaps shared with other models. This multiple interaction poses some construction decisions about the simulation.

One approach would be to construct the simulation as a monolithic structure. All the models are build and interlocked in the overall simulation. As the simulation proceeds, no model can inadvertently alter data that might be used by another model. This is not a very flexible design, but is used quite extensively.

A better approach is to use modular construction with each model being designed for optimum efficiency. As long as the models adhere to certain protocols, they can interact with each other. This is a very difficult architecture to achieve but most modelers are pursuing this strategy (Costanza th proceeding,Maxwell and Costanza, this proceeding).

Current Approaches

How can we incorporate multi-resolution data into our dynamic simulations? Let's take a look at some possible solutions to the two main pieces: asynchronous time and multiple spatial resolution.

In the research at the University of Arizona, we have been pursuing the use of the DEVS formalism. DEVS was originally implemented in Scheme, but the overhead imposed by Smalltalk limits its use for very large scale simulations. A new version of DEVS was written using C++ and has proven to be quite portable and computationally efficient. Additional refinements such as quantizing the events and the time have increased the efficiency of large simulations by several orders of magnitude with no loss in the accuracy of the simulations. The inherent properties of DEVS gives us asynchronous timing without additional overhead. Each part of the simulation proceeds at its appropriate time.

The standard discrete event approach imposes some restrictions on how the event cycles is formulated. One of the current directions being pursued by our team is the use of quantization of events and time. This techniques makes the steps between events smoother and helps to increase the efficiency of the simulation. Preliminary results show that we can speed up the simulation by an order of magnitude with no loss of information.

Another aspect being examined is the use of post processing to add continuous information back into the discrete time steps. This would allow us to improve specific components of the data stream when necessary. This could be very important for visualization techniques.

The problems of data handling pose a more difficult problem. Let's asume that the simulations are driven using GIS databases. Most models designed to use GIS are specific to a particular GIS program or use some form of export file to access data. To keep data preparation and file structures to a minimum, we opted for an approach that permitted model driven design.

The implementation of a multi-resolution data scheme into a simulation using a GIS database requires a fundamental understanding of what models do with data. At the basic level a model performs either an IO operation or a calculation. Since the calculation is irrelevant to where the data came from or where it will go, the only real problem then is the IO step.

All GIS databases have a data structure which is usually not easily accessed. Therefore options such as common data transfer formats have been getting a lot of press. This is still an export strategy and not very efficient. The best approach is to make use of the georeferencing of the data and have the model simply request what it needs based on a coordinate pair. This immediately does two things.

First, the model does not have to be designed for a specific GIS database. Since all GIS databases can respond to a coordinate pair for information the procedure call becomes platform independent. Secondly, the information can theoretically be derived from either vector or raster data. What we have done in our research simulator is to use a C++ object that provides these capabilities.

The Common Object Data Interface is allows models of any type to access GIS databases. The current implementation is used only with raster data but the next generation will include vector data access. The object design can be thought of as an interface between the data and the models. A model asks the object to provide the necessary data and the object responds. If the model wants to store information the object can also perform that task.

The original object design was a single structure that all models accessed. As we moved to massively parallel computers and very large databases, the object was redesigned in a distributed version. With a fully distributed data handler it is possible to simultaneously provide data at different resolutions to any number of models.

Visualization is a very important component in multi-resoluton simulation. The compound problems of asynchronous models and different data resolution requires a better approach to visual display than what is currently available. For example, we want to be able to examine what is happening in any part of the simulation while the program is still running. This entails developing new software to integrate the appropriate data streams and then display them in perspective rendering. The current prototype implementation of this software runs on an SGI machine. We are currently making a port to an X-based display.

Handling all the data traffic for this type of simulation also requires better communication between software and hardware. We are working in a heterogeneous environment comprised of Unix workstations and supercomputers. To handle the requests for data and control of the distributed simulation we have been using the Schooner interconnection system. This software allows programs on different machines to communicate regardless of the operating system, hardware architecture, or programming language used. We have successfully linked SGI and SUN workstations to run a simulation and move the visualization data stream from the SUN to the SGI in real time.

Conclusion

The use of multi-resolution data in large scale simulation has meant abandoning the traditional modeling concepts. As the benefits of GIS databases were recognized, modelers tried to fit the models to the data. In so doing, they kept the model from doing what it should, which are calculations.

If the modeling community is going to make the best use of the data in the GIS systems, we have to acknowledge what GIS does best, which is not dynamic modeling. Modeling needs to use new approaches.

Acknowledgments

This project is funded in part by NSF HPCC Grant ASC-9318169, and NSF Grant ASC-9204021.

References

Costanza, Robert, 1996. The future of spatial modeling for understanding and predicting landscape transformations. This Proceedings.

Maxwell, Thomas and Robert Costanza, 1996. Distributed Modular Spatial Ecosystem Modelling. This Proceedings.

Zeigler, Bernard P. 1976. Theory of Modelling and Simulation. John Wiley, New York.

Author Information

George L. Ball, Asst. Research Professor, School of Renewable Natural Resources, University of Arizona, Tucson, Arizona 85721, Phone: (520) 621-5951, FAX: (520) 621-8801, Email: gball@nexus.srnr.arizona.edu

Bernard P. Zeigler, Professor, Electrical and Computer Engineering, University of Arizona, Tucson, Arizona 85721, Phone: (520) 621-2801, FAX: (520) 621-8076, Email: zeigler@ece.arizona.edu

Richard Schlichting, Professor, Computer Sciences, University of Arizona, Tucson, Arizona 85721, Phone: (520) 621-4324, FAX: (520) 621-4246, Email: rick@cs.arizona.edu

Michael Marefat, Asst. Professor, Electrical and Computer Engineering, University of Arizona, Tucson, Arizona 85721, Phone: (520) 621-4852, FAX: (520) 621-8076, Email: marefat@ece.arizona.edu

D. Phillip Guertin, Assoc. Professor, School of Renewable Natural Resources, University of Arizona, Tucson, Arizona 85721, Phone: (520) 621-1723, FAX: (520) 621-8801, Email: phil@nexus.srnr.arizona.edu

Gary D. Bishop and M. Robbins Church

PRODUCTION OF REGIONAL MAPS OF LONG-TERM RUNOFF USING SIMPLE GIS-BASED METHODS

We conducted research to develop and test the accuracy of nine simple GIS-based methods for producing long-term runoff contour maps. Our goal was to create maps as accurate as those produced with manual methods by the U.S. Geological Survey (USGS). One of our methods uses available gaged runoff data only. Four use available gaged runoff data and measured precipitation. The remaining four methods use gaged runoff data and precipitation estimates from PRISM (Precipitation-elevation Regressions on Independent Slopes Model).

We evaluated our maps with both qualitative and quantitative methods. These methods included a visual comparison of the GIS-based maps with a USGS manually produced map and a statistical comparison of runoff estimated at precipitation stations in our study region with values interpolated from the manual map. We also conducted an uncertainty analysis in which we compared deviations of computed (interpolated) runoff from actual gaged values at 93 sites. Overall, the automated methods performed as well as, or better than, the manual method. Based on our evaluations we feel that the method using PRISM-based runoff-to-precipitation ratios gives the best results. Our work indicates that simple GIS-based methods can produce long-term runoff contour maps with regional accuracies equivalent or superior to those produced by manual methods.

Gary D. Bishop
Ogden Professional Services
U.S. EPA - NHEERL
Corvallis, OR 97333 USA
(541) 754-4807, FAX (541) 754-4716
E-mail: bishop@mail.cor.epa.gov

and

M. Robbins Church
U.S. Environmental Protection Agency
National Health and Environmental Effects Research Laboratory
Corvallis, OR 97333 USA
(541) 754-4424
E-mail:church@mail.cor.epa.gov

W.G. Booty, D.C.L Lam, D.A. Swayne, C.I. Mayfield, L. Leon Vizcaino, G.S. Bowen, I. Wong, D.F. Kay

SYSTEM INTEGRATION OF GIS AND ENVIRONMENTAL MODELS IN THE PERSONAL COMPUTER ENVIRONMENT

ABSTRACT

We will demonstrate progress in developing multi-station applications for environmental modelling using Geographical Informations Systems tools in conjunction with engineering models. We have overcome many the problems in developing multi-function systems with a spatial processing dimension, using the UNIX, Windows*, Windows '95*, and Windows NT* plaftorms. Our experiences have included developing client server applications (using several database products including Oracle, and with models for source, transport and fate of aquatic contaminants such as AGNPS). As well, we have incorporated models for the effects of contamination on existing flora and fauna, and for changes in the ecosystems wrought by human activity. The whole system has a common "look and feel". It makes an extensive use of "wizards" for reducing brittleness and improving user satisfaction and productivity.

In this poster / demonstration display, we show the conceptual framework of the integrated system, acronymed RAISON. For input, it uses numeric data, textual descriptions or meta-files, GIS maps, photos and satellite images, as well as linkages to models. It offers a generic framework to apply these types of inputs for problem-solving and decision support, using a variety of software functions and tools, including statistical analysis, expert systems and neural networks, to produce customized interfaces and output such as interpretation, advice, scenario tests and strategic and policy recommendations. Models can be incorporated into the system by: (a) using the codes as given, (b) building an interface that intercepts the model input and output files by running the model off-line, or (c) emulating the model by a simpler input-output or reduced form model.

An example used in the demonstration is the linkage of the AGNPS model developed by the U.S. Department of Agriculture for analyzing non-point source pollution due to rainfall events over an agricultural watershed or basin. By using GIS layers as objects and overlaying model grid cells on them, we show how information is transferred and helps to define model coefficients. The use of map-based and keyword-based queries also helps to retrieve appropriate information for direct and indirect input to the model. These technologies can be adapted for other problems. One example is the application of point-source models for industrial or sewage effluent. Environmental conditions can be defined by statistically ranking observed downstream water quality data using map-based queries. Confirmation of possible

impact can also be tested with 1- or 2-dimensional water quality models using various scenario inputs.

We have also begun extensive development in World-Wide Web applications for environmental management, for example in our agricultural organic fertilizer management program resident on our web site HTTP://ozone.crle.uoguelph.ca/manure/. We will here, or in a separate paper, discuss the difficulties and promise in WWW applications development. For instance, a common interface with a client-server architecture reduces the burden of multi-platform development. At the same time, the dependence on common-carrier network poses risks for the production user of computing resources.

* copyright Microsoft Corp.

W.G. Booty, National Water Research Institute, Environment Canada
Burlington, Ontario, Canada, L7R 4A6 (email:bill.booty@cciw.ca)

D.C.L. Lam, National Water Research Institute, Environment Canada,
Burlington, Ontario, Canada, L7R 4A6 (email:david.lam@cciw.ca)

D.A.Swayne, Computing and Information Science, University of Guelph
Guelph, N1G 2W1, Canada (email: dswayne@uoguelph.ca)

C.I. Mayfield, Environmental Biology,University of Waterloo,
Waterloo, N2L 3G1, Canada (email: mayfield@sciborg.uwaterloo.ca)

L. Leon Vizcaino, Civil Engineering, University of Waterloo,
Waterloo, N2L 3G1, Canada (email: LFLEONVI@uwaterloo.ca)

G.S.Bowen, Ontario Ministry of Environment and Energy,
Toronto, Ontario, Canada M4V 1P5 (email: boweng@gov.on.ca)

I.Wong, National Water Research Institute, Environment Canada,
Burlington, Ontario, Canada, L7R 4A6 (email:isaac.wong@cciw.ca)

D.F. Kay, National Water Research Institute, Environment Canada,
Burlington, Ontario, Canada, L7R 4A6 (email:doug.kay@cciw.ca)

Patrick J. Bresnahan*, David J. Cowen, John R. Jensen, and
Halkard E. Mackey, Jr.

SPATIAL MODELS HIGHLIGHT RADIOLOGICAL HAZARDS

As environmental characterization and remediation activities proceed at Department of Energy (DOE) facilities, radiological contamination of fluvial systems is becoming more important in overall risk assessments. The application of GIS techniques are expanding to coordinate and analyze the variety of sample data needed to define current and potential hazards. Previous research has concentrated on independent groundwater, surface water, and air dispersion modeling related to chemical and radiological components. In moving toward more comprehensive analysis of multiple exposure pathways, aerial gamma data has been converted into various GIS products as additional inputs to such assessments.

Radiological contamination at the Savannah River Site (SRS) is concentrated around production facilities and in fluvial sediments and associated vegetation. Spatial delineation of radiological hazards in stream corridors has proven difficult due to numerous assumptions required to quantify contributions from discrete elements. Furthermore, sampling programs are often limited in areas inaccessible to survey crews. Aerial gamma surveys have been used to provide an overall picture of radionuclide distribution in the environment. By incorporating aerial gamma surveys, as well as ground-based sampling, in a structural model, it has been possible to convey the extent of radiological hazards. This poster displays the representation of remotely sensed gamma data in surface models to produce potential hazardous waste unit boundaries to be used in guiding site-specific surveys.

From the 450,000 individual gamma response samples recorded for each radioisotope, a point coverage was created. Before the data were used in any analyses, the spatial and spectral resolution of the aerial platform was assessed. Regression analyses were conducted to confirm the integrity of the data by comparing in-situ gamma measures and aerial samples. Triangulated irregular networks (TINs) were created from these point layers to be used in several analytic operations. The primary goal of this research was to use the TIN models as the source from which contour boundaries could be derived for specific radioisotopes. The resulting procedures allow the definition of user-defined boundary conditions for discrete exposure levels.

Beyond defining contoured boundaries of radioactive contamination, the TINs were used in several visual modeling projects to convey multi-dimensional relationships between hazards and environmental features. Methods for interpolating exponential data and representing uncertainty in radiological assessments were also investigated. Additional information regarding the spatial distribution of radioactive contaminants was gained from the classification of TIN slope. This measure distinguishes point source elements from homogeneous distributions of radioactivity.

As risk assessments begin to develop a synoptic view of hazards, it is hoped that this research will allow radiation information to be incorporated into predictive models and comprehensive risk assessments.

Patrick J. Bresnahan*, David J. Cowen and John R. Jensen
Department of Geography
University of South Carolina
Columbia, SC 29208
(803) 777-4590, FAX (803) 777-7489
E-mail: pat@otis.cla.sc.edu
cowen@otis.cla.sc.edu
johnj@kirk.geog.scarolina.edu

Halkard E. Mackey, Jr.
Westinghouse Savannah River Company
(803) 725-5322

---

* This research was supported, in part, by an appointment to the U.S. Department of Energy Laboratory Cooperative Research Training Program at the Savannah River Technology Center administered by the Oak Ridge Institute for Science and Education.

Additional support was received from the South Carolina Universities Research and Education Foundation (SCUREF), Task F111.

Douglas Briggs, Jim Westervelt, Shaun Levi, Steve Harper

A Desert Tortoise Spatially Explicit Population Model

Introduction

Across the nation, land management offices responsible for the management of natural resources, endangered species, water quality, aesthetics, and economic productivity are turning toward **dynamic spatial ecological modeling**. Military land managers, continuously facing the challenge of accommodating both the legislated conservation measures such as the Environmental Protection Act, designed to protect ecosystems on military installations, and the demands of rigorous training schedules, can use dynamic spatial ecological models to investigate the impact that various training schedules and other human impacts have on the environment in both the short term and the long term. With the use of the modeling tools generated by this research, military land managers can develop a training schedule that both satisfies training requirements and preserves the ecological systems within which the training will occur.

A dynamic spatial ecological model is a computer model that evolves in simulated time. Research teams create and develop the complex system of mathematical, logical, and **stochastic processes** that govern the progress of the **computer simulation**. A set of initial conditions (culled from raster images, vector, or survey data, for example) seeds the system and provides initial conditions for the simulation to develop in time.

This research project developed a prototype dynamic spatial ecological model of the **endangered desert tortoise** *(Gopherus agassizii)* population on the grounds of the Fort Irwin Army Training Center in the Central Mojave Desert of California. A 57-by-57 grid of 1 kilometer square areas divided a sample portion of the landscape of Fort Irwin into "cells." Each cell is assigned an identical computer model that simulated pertinent environmental variables such as elevation, soil type and moisture, precipitation, air and surface temperature, tortoise mortality, and percentage of vegetation cover. The model itself, developed in four parts by independent, **interdisciplinary research teams**, was written with the **STELLA modeling software** and captures the hydrology, vegetation cover, tortoise population, and tortoise migration dynamics as they evolve in monthly time-steps. This poster describes the process by which the model was developed and executed as a part of the larger simulation of the entire Fort Irwin ecological landscape.

The Spatial Modeling Environment

After the STELLA model is complete, it is converted into the computer simulation which will execute a copy of the model in each of the cells simultaneously and in parallel. We used Dr.

Thomas Maxwell's Spatial Modeling Environment (SME) software tool for this conversion process.

First, the STELLA model is saved as a text file which summarizes the difference equations governing the model's stochastic processes. This text file is then translated in two stages: first into Modular Modeling Language (MML), and then into C++. **Raster GIS maps** for ground elevation, slope, aspect, vegetative ground cover, ground compaction, and soil water content are linked to the SME model for **initialization data**, **output preferences** are registered in a configuration file, and the C++ code is linked and compiled. We then run the resulting **executable binary** on a UNIX workstation and both observe the output as the simulation runs, and collect it for further analysis later.

Future Directions

The Desert Tortoise Model demonstrates how dynamic, spatial, ecological modeling tools can be designed for effective use in land management. Though similar in some ways to the Sage Grouse model (Westervelt, 1995) the Desert Tortoise model pushes the frontiers of these research efforts by implementing a land management tool, as opposed to a demonstration device. The Desert Tortoise model presented here is a broad-based picture of the relevant dynamics of tortoise habitats at Fort Irwin. In all cases, a model is a simplification of the system that it represents and for that reason this model provides most accurately predictive results when run for no longer than 100 years. For this first stage of development the critical components of development included the multidisciplinary research team; a coherent, logical.modeling process; and a collection of modeling software tools capable of managing the type and volume of data needed to effectively model a multi-cellular landscape. The software tools we used were:

- GRASS -- The Geographical Resources Analysis Support System, which helped initialize the model.
- STELLA -- A graphical user interface oriented dynamic programming language that enabled all researchers regardless of modeling experience to participate in the simulation model design process.
- SME -- The Spatial Modeling Environment tool designed by Dr. Thomas Maxwell of the Univeristy of Maryland's International Institute of Ecological Economics converted STELLA models into a form that could be at a multi-cellular level on desktop workstations.

The Desert Tortoise Model research efforts to implement dynamic, spatial, ecological models as effective landscape management tools are not yet finished. The first phase comprising the development of the spatial model is complete. We now look to develop a battery of powerful yet comprehensible statistical tests for performing sensitivity analyses to validate the model and its output. We plan also to continue to integrate into the STELLA model our expanding empirical knowledge of tortoises and their habitats, in order to improve the predictive power of the model.

Authors and Acknowledgements

Douglas R. Briggs, Dept. of Computer Science, University of Illinois at Urbana-Champaign. (drbriggs@cs.uiuc.edu)

Dr. James Westervelt, Dept. of Urban and Regional Planning, University of Illinois at Urbana-Champaign. (westerve@gis.uiuc.edu)

Shaun Levi, Dept. of Geography, University of Illinois at Urbana-Champaign. (s-levi@uiuc.edu)

Steve Harper, Dept. of Ecology, Ethology, and Evolution, University of Illinois at Urbana-Champaign. (sjharper@uiuc.edu)

The authors gratefully acknowledge the work of Dr. Thomas Maxwell of the Chesapeake Biological Laboratory at the University of Maryland for his development of the SME software. This study was conducted for Headquarters, US Army Corps of Engineers under Project 4A1102AT25, "Environmental Restoration", Work Unit IA4, "Fundamentals in Dynamic Ecological Modeling".

Further Information

More information about this and related projects at the Geographic Modeling Systems Laboratory at the University of Illinois at Urbana-Champaign/U.S. Army Construction Engineering Research Laboratory is available here.

Greg Cole

GENERATION AND UTILIZATION OF DEM DATA FOR ENVIRONMENTAL RESTORATION - LOS ALAMOS NATIONAL LABORATORY

A high-resolution, aerial survey was flown in 1991 to support environmental restoration (ER) activities at Los Alamos National Laboratory (LANL). Color aerial photos and orthophotos, digital contour, and other data were collected by this survey. A digital elevation model (DEM) was created from the contour data at a resolution of 1-foot. Software was developed to create the gridded elevations through a combination of interpolation and splining techniques. Software development and generation of two (2) billion gridded elevations were accomplished in a 4-month period, utilizing an HP workstation. The resultant data are utilized in combination with digital orthophoto data to provide geographic visualization of ER activities and data.

Detailed Author Information

G. Cole, EES-1, MS D462, Los Alamos National Laboratory, Los Alamos, NM 87545 - phone: 505-667-1858, fax: 505-665-3285, e-mail: gcole@lanl.gov

Dwayne E. Porter[1,2], David J. Cowen[2], Don Edwards[1,3], Tom Siewicki[4], Ben Jones[1] and Christopher Corbett[2]

INTEGRATED RISK AND IMPACT ASSESSMENT MODELING IN SUPPORT OF SUSTAINABLE DEVELOPMENT IN AN ESTUARINE ENVIRONMENT

Abstract

Problems associated with developing integrated environmental models for sustainable development range from understanding the molecular/physiological mechanisms of toxic effects on organisms to understanding the physical/chemical mechanisms of transport and fate of pollutants, nutrients and water to validating predictive models of how ecosystems respond to developmental pressures. Fortunately, the advancement of Geographic Information Systems (GIS), remote sensing, Global Positioning Systems (GPS) and environmental statistics are enabling researchers and resource managers to develop techniques previously too complex or disparate to quantify. The complexity and severity of environmental impacts associated with coastal development necessitates resource managers to explore new spatial analytical techniques combined with multi-disciplinary scientific expertise for proactive coastal zone management.

Arising from these environmental concerns and the identified need for better databases and integrated spatial models for sustainable development, a long-term study of the impacts of urbanization on localized coastal estuaries of the southeastern United States was initiated in 1990. An overall goal of the study is the utilization of the tools of Geographic Information Processing (GIP) to integrate the necessary data and scientific expertise for the identification, assessment and modeling of relationships within coastal estuaries and impacts associated with anthropogenic and physiographic activities. This goal is being achieved through the implementation of a multi-participant GIS, and the development and validation of integrated spatial models. This work presents spatial models which assess land use and land cover characteristics to predict the impacts of urbanization on critical estuarine habitat and natant fauna. In addition, a working concept is described for the assimilation and dissemination of data, metadata and derived information necessary for integrated, multi-disciplinary environmental research. This approach can be used by coastal resource managers to predict the impacts of proposed landscape modifications prior to any changes taking place.

---

[1]Belle W. Baruch Institute for Marine Biology and Coastal Research, and the Marine Science

Program2
University of South Carolina
Columbia, SC 29208
803-777-4615 (P) 803-777-3935 (F)
dporter@hugo.geol.sc.edu (E-mail)
2Liberal Arts Computing Lab and Department of Geography
University of South Carolina
Columbia, SC 29208
3Department of Statistics
University of South Carolina
Columbia, SC 29208
4NOAA National Marine Fisheries Service
Southeastern Fisheries Science Center - Charleston Laboratory
Charleston, SC 29208

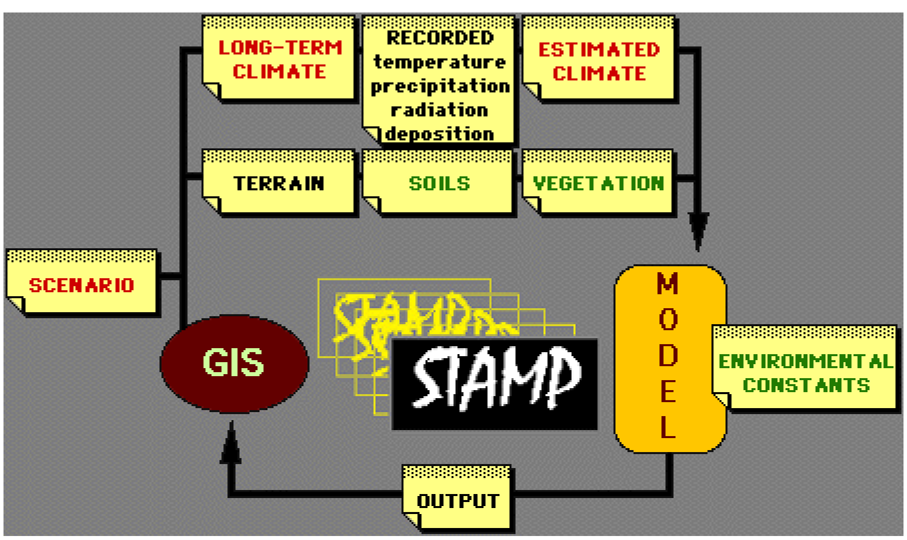
SPATIO-TEMPORAL OBJECT HANDLING FOR MODELING THE EFFECTS OF ACID DEPOSITION

Ferko Csillag, Scott Mitchell, Rebecca Handcock
Department of Geography, University of Toronto
Charles Driscoll, Kiran Sequeira
Department of Civil and Environmental Engineering, Syracuse University
Hon Yau
Northeast Parallel Architecture Center, Syracuse University
http://eratos.erin.utoronto.ca/fcs/intro_fcs.html fcs@geog.utoronto.ca

---

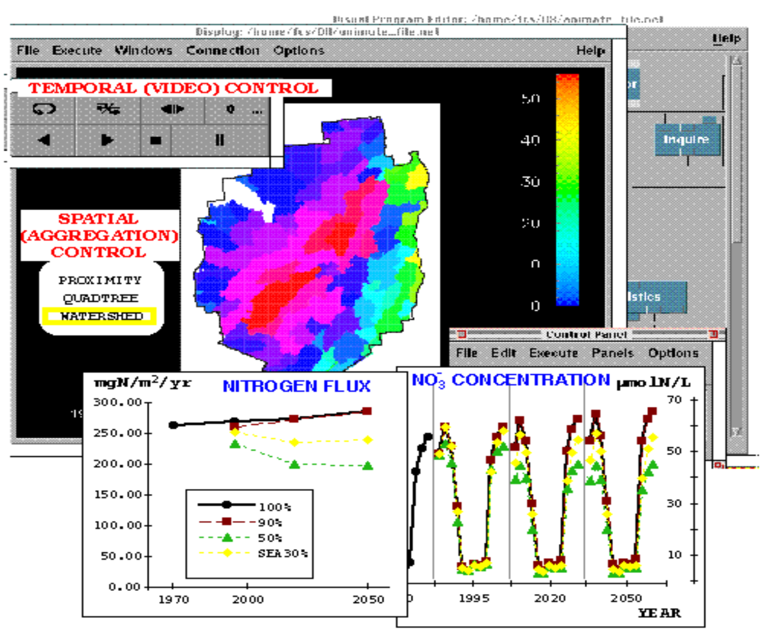
ABSTRACT

Modeling the long-term effects of acid deposition on forest and aquatic ecosystems under various scenarios in the northeastern US requires the integration of a variety of **data** and **software tools** . We have developed an integrated system for the spatially explicit parametrization of PnET (*Aber and Federer 1992*). The integration is based on object-oriented extensions to ArcInfo. The **model-object** is a concept that incorporates **spatial and temporal** constraints of the environmental model in the GIS interface. In ***STAMP***, this loosely-coupled system of GIS, environmental model and visualization tools, for example, animating/displaying time-series of (climate) data, aggregation of variability by watersheds or by proximity polygons or by adjusting partitions to local heterogeneity is straightforward (*Csillag et al. 1995*). The approach, which incorporates measures of variability, significantly reduces the problems of data heterogeneity in both spatial and temporal context (e.g., intensively studied sites versus relatively unknown areas), and easily supports sensitivity analysis.



To apply PnET at regional scales it is necessary to estimate long-term climate and deposition conditions where they are not available. The current version is set up in such a way that the site-based bucket model is "extended" to a neighborhood according to some partitioning criterion. The spatial heterogeneity of each "modeling unit" is computed and stored within the GIS and are used in parameterizing the model. The model always looks for measured and/or mapped data of temperature, precipitation, solar radiation, wet and dry deposition, soil properties (e.g. water-holding capacity) and vegetation characteristics (e.g. leaf N concentration) substitutes estimates when encountering missing data.

Our test-area, the Adirondack Mountains in northern New York, receives elevated inputs of precipitation and acidic deposition (SO4 deposition is approximately 500 eq/ha-yr, NO3 deposition is about 400 eq/ha-yr). The combination of elevated atmospheric deposition of strong acids and edaphic/geologic characteristics places terrestrial and aquatic resources at risk. Comprehensive lake surveys and long-term monitoring has revealed that a significant proportion of lakes are impacted. In spite of decreasing atmospheric inputs, the acid neutralizing capacity (ANC) of most lakes has not increased. Decrease in ANC appears to be due to increases in NO3 concentrations, which appears to be due to decreases in the retention of N by older stands of forest vegetation (*Driscoll and VanDreason 1993*). Using various deposition scenarios, we have identified risk levels of nitrogen saturation along a significant gradient of environmental conditions in the Adirondack Mountains (e.g., topography, deposition, geology, vegetation).



This image illustrates the coupling of GIS and PnET-model via spatio-temporal object handling and visualization. Spatial partitions are created and stored by the GIS; PnET is parameterized and run on these spatial units predicting, in this case, N-budget. The (monthly) output can be "played" on a video controller and/or aggregated within the GIS. The image of the Adirondack Park displays by second-level watersheds annual N-flux (loss). The high values correspond to higher N-input (primarily at higher elevations). Uncertainty and bias in the output is a function of spatial partitioning. The bottom figures illustrate four different scenarios for N-input. 100% stands for "business as usual", 90% and 50% mean respective amounts of 1980 values kept constant in the future, and SEA_30% stands for seasonal 70% reduction of N-input. With the first two scenarios "on average" N-saturation would be reached in 200-300 years; with the latter two scenarios we found siginifcant decrease in N-loss to surface waters.

Our immediate plan is to combine PnET with CHESS (*Santore and Driscoll 1995*). The linked model will be able to provide a more realistic prediction of biogeochemical processes. We are also testing a more modular setup of the model, in which various partitions can be combined with each other. It is also anticipated that long-term scenarios will be analyzed for the entire NE-American continent.

This project is supported by IBM, NSF, EPA and NSERC.

REFERENCES:

- Aber, J. and Federer, A. (1992) A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. Oecologia 92:463-474.
- Csillag, F. (1995) On uncertainty in geographic databases driving regional biogeochemical models. AutoCarto-12, ACSM/ASPRS, Bethesda, MD., pp.264-271.
- Driscoll, C. and VanDreason, R. (1993) Seasonal and long-term temporal patterns in the chemistry of Adirondack lakes. Water, Air and Soil Pollution 67: 319-344.
- Santore, R. and Driscoll, C. (1995) The CHESS model for calculating chemical equilibria in soils and solutions. In: R. Loeppert et al. (eds.) Chemical Equilibrium and Reacction Models. Agronomy Society of America, Madison, WI.

CHARACTERIZATION OF THE RECHARGE AND DISCHARGE COMPONENTS OF THE DEATH VALLEY REGIONAL GROUND-WATER FLOW SYSTEM USING REMOTE SENSING AND GIS TECHNIQUES

by
Frank A. D'Agnese, Claudia C. Faunt, and A. Keith Turner

Remote sensing and GIS techniques were utilized to develop maps that described recharge and discharge components as part of continuing comprehensive regional hydrologic modeling studies of the Death Valley regional ground-water flow system. These techniques allowed the integration of disparate data types from various disciplines, including hydrogeology, soil science, and plant ecology, to develop a spatially complex representation of surface and subsurface hydrologic processes. Multispectral remote sensing data was evaluated using image classification methods to produce a vegetative cover thematic layer, that provided data for estimating rates of both evapotranspiration and infiltration.

Difficulties in discriminating vegetation in arid regions from remote sensing data were resolved, and a method was developed that converted land-surface spectral responses into vegetation community classes. The resulting vegetation-landform map described eleven major classes. It formed the basis for subsequent evapotranspiration and infiltration estimations.

Previous evapotranspiration studies in this region included empirical, water-consumption estimates for phreatophytes. The vegetation map and ancillary data sets were combined in a GIS to delineate previously disregarded evapotranspiration sources for wetland, shrubby phreatophyte, and wet playa areas. Specific water-use rates, described by data developed from independent field studies, were then applied to mapped evapotranspiration areas and volumetric evapotranspiration discharge rates were calculated.

Ground-water recharge estimates were developed by refining an empirical, area-altitude method developed by Eakin and others in 1951. Preliminary recharge estimates were defined using this method, but were refined and modified by incorporating data that described varying soil moisture conditions. Additional potential recharge indicators were identified: (1) topography (elevation), (2) slope aspect, (3) parent material (rock and soil permeability), and (4) vegetation zones. GIS methods were used to combine geologic and topographic data with the vegetation-landform map to produce a map describing recharge potential on a relative scale. The resulting recharge potential map indicated areas within the Death Valley region that possess high, high-moderate, moderate, moderate-low, low, and no potential for regional ground-water recharge. These classes were then assigned percentages of average annual precipitation that would be contributed to ground-water recharge in a manner similar to that used previously by Eakin and others. Improved recharge estimates resulted from the replacement of Eakin's original area-altitude classes by these comparative recharge potential classes.

The discharge and recharge maps were in raster formats and so were readily converted to data arrays defining initial rates of flux into and out of the ground-water system and incorporated into a regional finite-difference ground-water flow model. The detailed maps also were useful in calibration because they helped to describe the spatial heterogeneity of otherwise lumped model inputs.

Quality Assurance in GIS and environmental modelling, applied to an example of phosphate-saturated soils

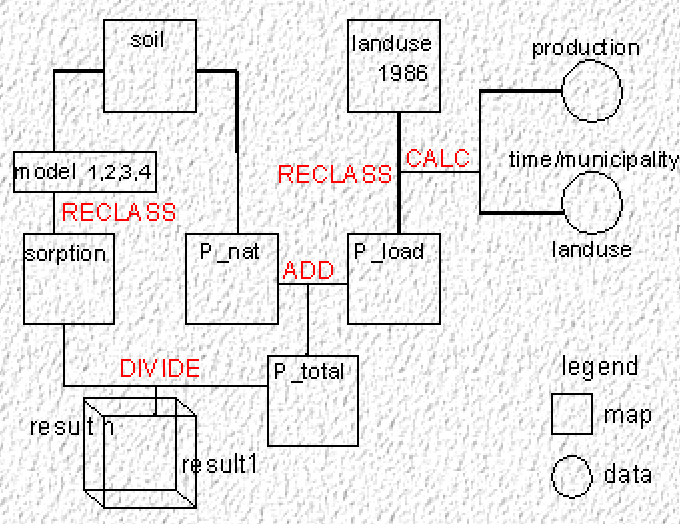
---

Introduction

The increase in use of GIS and environmental modelling and the availibility of many digital datasets raises the question about the quality of the end-products.
Especially for the layman end-user a procedure should be developed to qualify and guarantee the result. If a end-product will have its own 'life', outside the direct influence of the developer, it should satisfy strict quality standards.
Recently many authors (e.g. Goodchild 1993) discussed quality aspects of the data, procedure, visualization, error propagation and reliability of the model, but rather in isolation of the end-user. The combined effects of these quality aspects on a typical GIS application have, as far as reported in literature, not yet been discussed.

Application

In order to check the effects of an integration of the above mentioned aspects, a typical environmental pollution problem, i.e. phosphate-saturated soils in the Netherlands is studied (De Bakker et al 1995).
Soil data (1 : 50,000) are reclassified to indicate the natural phosphate amount (**P_nat**) in 1950 and the phosphate sorption capacity (**sorption**).
Agricultural statistical data on municipal level for several years are used to indicate manure production and changes in landuse. The calculated phosphate production is combined with landuse/cover derived from remote sensing images to map phosphate load (**P_load**) as a function of landuse and production. The phosphate content (**P_total**) as **P_load** and **P_nat** is divided by the sorption to indicate as result the phosphate saturated areas (see for procedure figure 1).



Assessment

Two methods of quality assessment are followed:

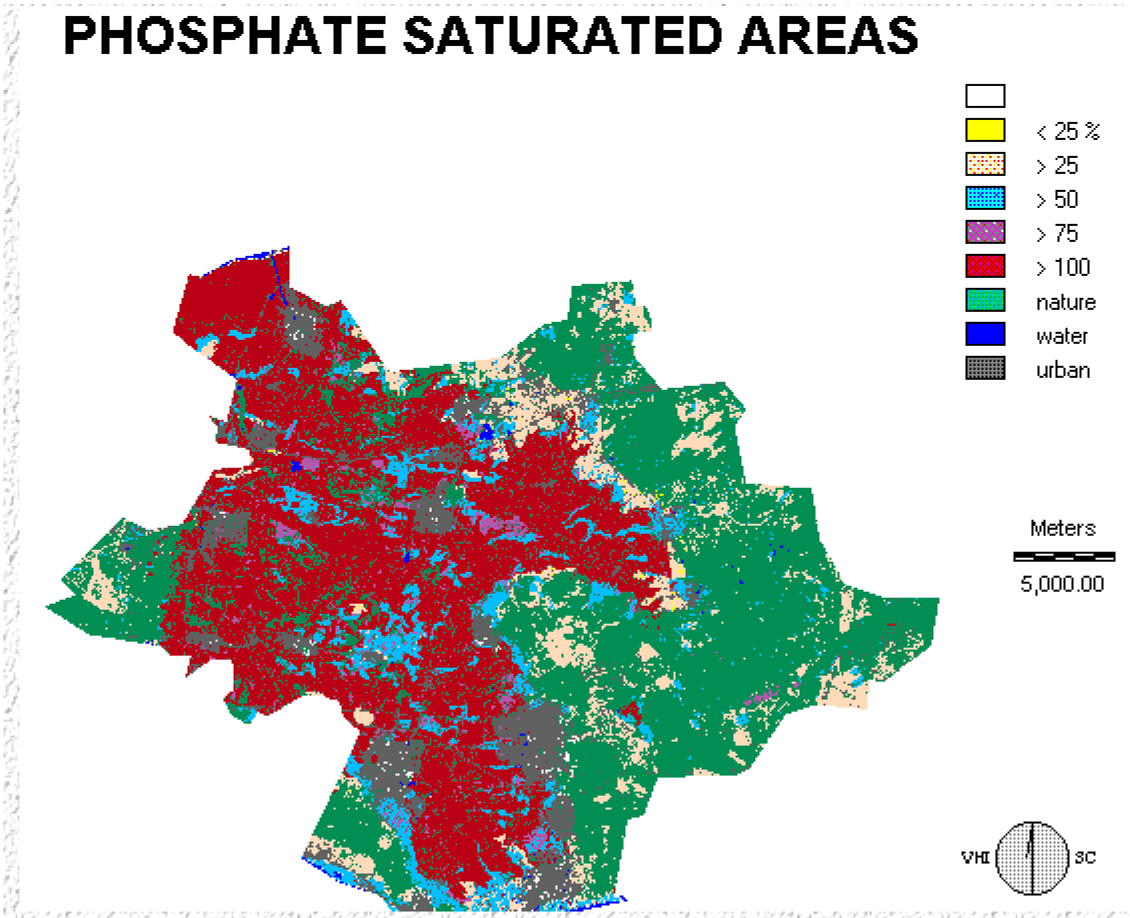
The first is a method of forward quality control. The used data and GIS procedure should be checked on errors (location and classification), error propagation, assumptions in the used models and final uncertainty. Quite often problems arise, because the used data have no attached meta-information. The interaction of locational and attribute error are not well indicated. By scenario-development and assumptions some insight can be given.

The second is a method of backward quality control. By interviewing the end-users a definition of the final result can be given. Questions like scale of publication, number of classes, colors of legend, number of maps, diagrams, charts, overall accuracy should be answered. The end-user defines the meta-information needed for a well balanced use.

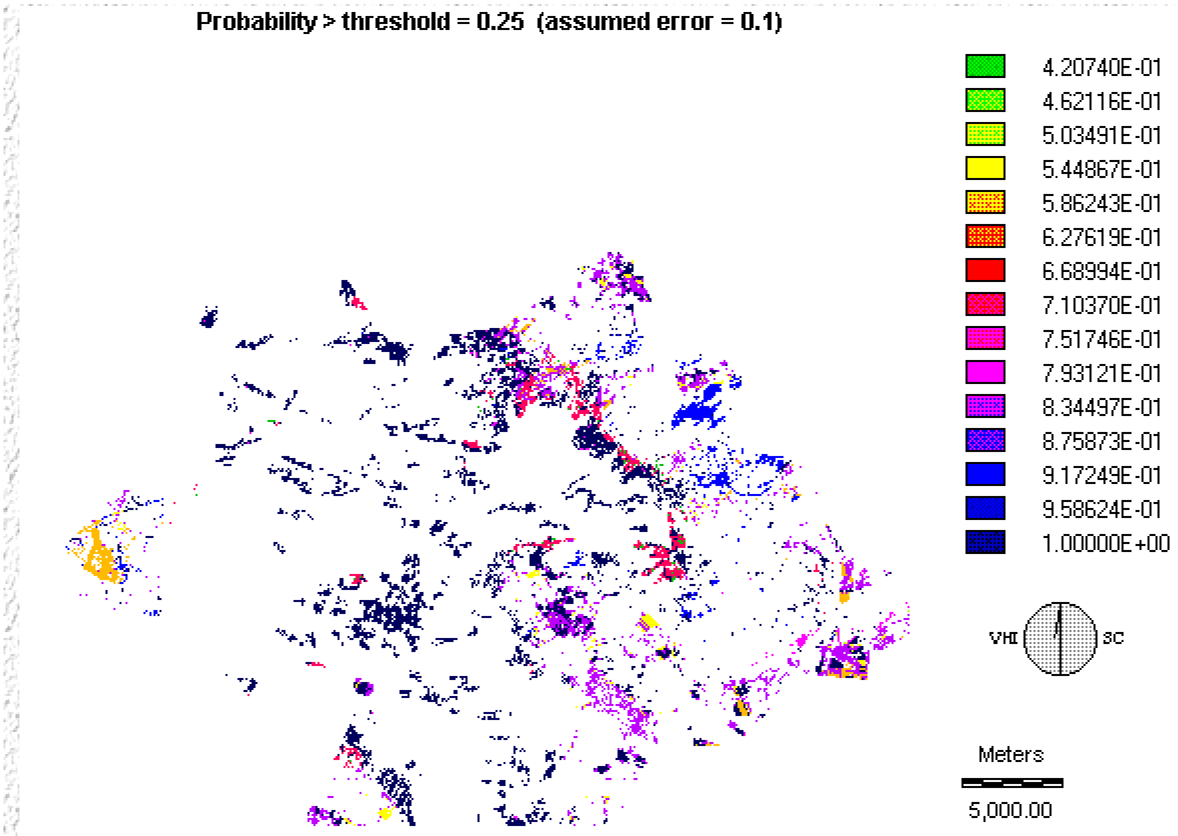
In reality most end-users have difficulties to answer these questions. Especially they like to have the best of everything.The combination and interaction of the two methods of quality assessment will show the end-user what he can expect of the used data, and if the final result will meet his standards.

Results

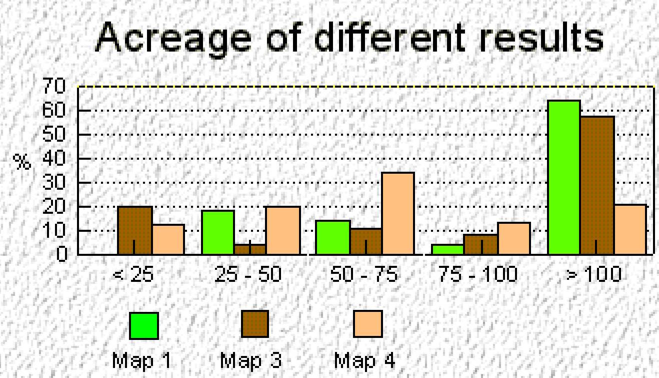
A possible final result (map 1) shows a worst case scenario (assumption: all produced manure is used inside the municipality).



Another result (map 2) indicates the possibility of exceeding the limit of 25 % of phosphate saturation (policy treshold) using an overall uncertainty of 10 %.



The upper and lower confidence limits using error propagation formulas are also results. The error or uncertainty can be different over space. Examples using an assumed standard deviation of 10 % (sd) are given according the following calculations:
(P_total - sd) / (sorption + sd) (map 3, not included) or **(P_total + sd) / (sorption - sd)** (map 4, not included).



In figure 2 the results of map 1, 3 and 4 are shown in the form of the difference in acreage over the legend classes.

Difficulties arise when the two assessment methods were compared. The confrontation with the uncertainty was not well recieved. Sometimes the end-users rejected the uncertainty in the results (and chose their political most satisfying option), or they rejected the total result (we cannot use this).

Conclusion

The application of the quality standards is depending strongly on the availibity of meta-information, knowledge of the developer how to use of meta-information and the insight of the end-user in the needed quality (required reliability level) of the end-product. For a robust backward quality assessment from end-user to developer some more testing of relevant questions and an increase of awareness is necessary. The possibilities and limitations in use of the GIS products will be then better indicated.

Acknowledgements

I like to thank the Winand Staring Centre, Wageningen for their help and the digital data. The SHAO is gratefully acknowledged for the subsidy of the development costs for the GIS application case for Higher Education (Bakker, de et al 1994). The system-managers of the Van Hall institute and Frits Steenhuisen are thanked for their support in "INTERNET" and "HTML".

Literature

1. Bakker, M. de, A.P.M. Valent & F.J. Blok, 1995, Incorporating spatial uncertainties in a phosphate saturation model, a GIS application in the Gelderse Vallei, the Netherlands, poster presentation: Joint European Conference and Exhibition on Geographical Information, The Hague, pp 466, 467
2. Bakker, M. de, F.J. Blok & J. Resink, 1994, Integration of GIS in regular education, an example in the environmental sciences, Proceedings fifth European GIS Conference, Paris, pp 262 - 268
3. Goodchild, M.F., 1993, Data models and data quality: problems and prospects In: Goodchild, M.F., B.O. Parks & L.T. Steyaert, Environmental modeling with GIS, Oxford University Press pp. 94 - 102

Marien de Bakker,
Van Hall instituut, Dep. of Environmental Sciences,
P.O. Box 17,
9700 AA GRONINGEN, The Netherlands
email: Marien.deBakker@vHall.nl

DIGITAL ELEVATION DATA AND GIS PROJECTS

Robert De Sawal
U.S.GeologicalSurvey
PO BOX 25046,MS-520
Denver,Colorado 80225
303-202-4804
fax 303-202-4504
e-mail rfdesawal@usgs.gov

ABSTRACT

The U.S. Geological Survey has been a producer and distributor of Digital Elevation Model (DEM) data since the 1970s. These data have been produced by photogrammetric methods and by scanning, tagging, vectorizing, and gridding of hypsographic map separates. Over several years a variety of gridding algorithms have been developed to produce elevation data from x,y,z point data. These algorithms influence the production of the elevation data. It is helpful, for successful GIS projects, if the user has some knowledge about the terrain and elevation data to insure that the ground resolution, horizontal accuracy, and vertical accuracy fit the project criteria. The future production of digital elevation data offers new methods of collection and terrain characterization. This poster session will demonstrate and discuss some of the methodologies and algorithms used in elevation data production, uses of elevation data, and future outlook of digital elevation data production.

METHODOLOGY OF DIGITAL ELEVATION PRODUCTION

Photogrammetric Methods

Digital ground elevation data is predominantly collected photogrammetrically, from stereo aerial photography, stereo satellite imagery, global positioning systems, ground surveying, and from scanned contour data vectorized as x,y,z point data.

One photogrammetric instrument, the Gestalt Photo Mapper 2 (GPM 2), was used by the U.S. Geological Survey (USGS) in the late 1970s and early 1980s to produce orthophotos and a magnetic tape recording of a Digital Elevation Model (DEM). These elevation data are a by-product of the stereo-correlation process. The data was recorded on magnetic tape, patch by patch, in a nominal 182 micron x 182 micron grid. The grid spacing within the patch is a hardware function and is the same regardless of the model scale. It is this patch quilt effect that may be detectable in 1970-1980 DEM data when displayed as shaded relief elevation data on a monitor. Most of the artifacts have been removed during editing prior to entry into the National Digital Cartographic Data Base (NDCDB) for archiving. Figure 1 Click Here for Figure depicts the USGeoData, Data Users Guides that provide information about Digital Line Graph and DEM data stored in the NDCDB. Quilt artifacts that do remain, generally are less then the required 15 meter maximum NDCDB elevation Root Mean Square Error (RMSE) required for GPM 2 data. However, depending on the application using these data the usual, 1 meter to 10 meter artifacts might cause a problem in the final results. Some of these data were able to meet the minimum 7 meter RMSE when regridded for 30-minute DEM products. For many projects these are not a problem but should be evaluated during project planning.

Another photogrammetric method of elevation production is the collection of elevation data by profiling directly from stereo aerial photography. Figure 2 Click Here for Figure and Figure 3 Click Here for Figure both show how this process is accomplished. At the USGS several of the DEMs in the NDCDB were produced using this method. Figure 4 Click Here for Figure shows what the photogrammetric profiling tracks look like in relation to the 1:24,000-scale topographic map, (1mm=78.74 feet). The aerial photography used was around 1:80,000-scale with a flying height of 40,000 feet. The physical stepover or distance between profiles on the imagery depended on the type of terrain encountered and typically ranged from 2 millimeters to 8 millimeters, but might have gone as high as 16 millimeters in very flat terrain. Figure 5 Click Here for Figure illustrates the potential for artifacts if the operator is fatigued, inexperienced or the collection rate along the profile is set to a rate that does not allow the operator to keep the floating dot on the model surface as viewed in 3-D stereo. Some of the differences are well within the 7 meter elevation requirements, but when viewed in stereo or if the contour interval is very small a linear pattern is easily seen. This may be important to be aware of when using these elevation data in a hydrologic GIS project.

Collection of specific terrain characteristics, such as ridge line, drain line and breakline data from photogrammetric means is a time consuming, but accurate way to input data into a Triangulated Irregular Network (TIN) method of terrain elevation production. The breakline collection insures that shore lines, cliffs, and man-made surface disturbances will be accounted for in the elevation generation. Figure 6 Click Here for Figure demonstrates how a portion of a stereomodel was collected using this methodology.

Scanned Vector Data Method

These data, sometimes referred to as, contour-to-grid, is a way to use the contour map information to produce gridded elevation data. The USGS, as well as the U.S. Forest Service, Natural Resources Conservation Service (formerly Soil Conservation Service) and some private companies, such as Infotec Inc., currently are using this method to produce digital elevation data. A film product made from the topographic map contour plate is scanned, usually at 500 or 1000 dots per inch, and the raster data entered into a software system that registers the data to a coordinate system, vectorizes the raster lines and interactively tag the contours, thereby assigning an elevation attribute to the linework. These lines are composed of x,y,z triplets, which will be input to a gridding program that will produce the digital elevation data. Depending on the algorithms used in the gridding process and smoothing ability of certain software products, artifacts can occur during the gridding. An example of a distance weighted algorithm artifact can be seen in the middle left portion of Figure 9. Click Here for Figure This usually occurs when there is a low density of point data, such as in flat areas where the contours are widely separated.

<u>Other Current Elevation Collection Methods</u>

The Global Positioning Systems (GPS) have become more widely used in collecting positional information about natural and man-made features. The vertical component is roughly half as accurate as the horizontal component, but for many projects would be more than adequate. The GPS hardware have become very reliable and portable allowing collection from a variety of means such as backpack, car, truck, dirt bike, train, etc..

Total ground survey stations are a very precise method for elevation data collection for small area projects or linear projects, such as pipe lines, transmission lines, and road construction. Generation of TIN elevation data from this collection process can yield very good results if significant changes in slope are captured.

Another method of collection is from stereo satellite digital imagery. Systeme Probatoire l'Observation de la Terre (SPOT), level 1A stereoscopic imagery is a source for the production of digital elevation data with greater than 10 meter vertical accuracy.

FACTORS AFFECTING DIGITAL ELEVATION DATA

Photogrammetrically profiled terrain digital data, as with other collection methods is dependent on the characteristics of the terrain. An in-house study at the USGS in 1978 reported that for 1:48,000-scale profiled stereo aerial photography resulted in sigma values of 10-15 feet for flat terrain, 10-25 feet for moderate hills and drains, 15-35 feet for mountain tops, and 15- 40 feet for steep sidehills. It has been observed that after many hours of collection some operators will exhibit fatigue by collecting digital elevation data lower than the terrain along a profile, and tend to float above the terrain along the adjacent return profile. This can be demonstrated from a contour plot, such as shown in Figure 5..

Producing a TIN from vector hypsographic data without breaklines can have an effect upon the formation of the triangles that compose the TIN. Drains do not form sharp patterns in the valley and tend to be smoother. Ridges tend to be slightly lower, unless an algorithm attempts to extrapolate beyond the last contour value at the top of the ridge, which might cause problems in discrepencies at the tops when trying to smooth. Shorelines are not clearly defined and blend with the ground and water. Without overedge data at map neatline edges void data tends to restrict the correct formation of the triangle at the edge. Figure 8 Click Here for Figure is an example of this problem when editing did not correct the problem. Incorporation of breaklines help to enforce the correct formation of the triangles and will produce very good results if the data being used is a good representation of the actual terrain and the significant changes in slope. Figure 7 Click Here for Figure demonstrates a TIN with only the use of hypsographic vector data.

Figure 10 Click Here for Figure illustrates with contours how lack of registration of drainage re-entrants with actual drainage, especially in mild- slope areas occur when only the hypsographic data is used in the production of the elevation data. Some algorithms will generate better surfaces with specific data types and all will generate good terrain surfaces when there is sufficient data that represents the significant changes in slope on the original surface.

SOME USES OF DIGITAL ELEVATION DATA

When a good digital representation of the terrain has been produced to meet the requirements of the project for which it was intended, then the use of these data in a GIS analysis will result in valid conclusions. When you know that the digital elevation model does not meet the needed requirements, but is the only data available then that will have to be conveyed to the customer of the GIS analysis in the final data to account for any discrepancy in the results.

An example of how digital elevation data is used in geophysical and exploration of mineral deposits is shown if Figure 11 Click Here for Figure and Figure 12 Click Here for Figure respectively. A point to keep in mind when incorporating elevation data with modeling algorithms is to have an understanding of the algorithm with respect to the elevation data available. If the elevation data used in your project is know to have deficiencies, then this information should be conveyed to the user to insure that the results are not being misinterpreted.

Visual portrayal of digital elevation data of the terrain is becoming a popular method in business, recreation, science, and military activities. As such, it is important to choose the best level of content for which the intended visualization is to be used. In the case of business locations, a coarse graphic representation is sufficient for a particular business theme. Other areas of business might want to portray demographic information on the digital elevation data to better convey the location of the information in 3-D representation. In the recreation area some companies of hiking trail maps want to show the terrain in a shaded relief representation derived from elevation data, and even overlay trails or imagery on the digital surface displayed in 3-D. Figure 13 Click Here for Figure is an example of digital shading created on a Mylar and overlaid on a 7.5-minute topographic quadrangle visually enhances the shape of the terrain. Many earth scientists are interested in the results of modeling the effects of their particular area of study with the digital elevation data. The military are active users of digital elevation data in tactical evaluations and training involving marine, air, and ground forces.

FUTURE TRENDS IN DIGITAL ELEVATION PRODUCTION

There are methods of digital elevation data collection that show promise for the future. The first is the autocorrelation and generation of elevation data from stereo digital imagery of any surface with identifiable control points. These data are produced from the softcopy photogrammetric workstation, which is currently becoming the next generation of photogrammetric workstation tool. Photogrammetry began with the analog plotter, then moved to analytical workstations, and now replacing the hardcopy stereo pair prints of imagery with the digital softcopy representation of a stereo pair of imagery. The digital elevation data produced using this technology represents a surface of natural and man-made features. For example, the elevation data will be autocorrelated on the tops of buildings, trees, towers, bridges, ground vegetation, etc.. Users of these data may find that by not having only the ground represented is a benefit in modelling activities such as: surface roughness in air flow over terrain, pollution tracking, commercial and military flying, tree heights for various deciduous and evergreen stands, line of sight for transmission signals, etc.. The need to have the elevation represent the ground level only involves editing of these data. Algorithms and methods to do this as an option is currently being worked on by vendors and academia.

Another automated method of elevation data collection for natural and man-made surfaces is from Interferometric Synthetic Aperture Radar (IFSAR). The positive aspect of this collection is the cloud penetration and limited surface penetration characteristics. The sensor can be located in high altitude aircraft and satellite systems allowing collection above storms and other flying aircraft. Choice of systems with specific band frequencies permit a sensor to collect information about surface characteristics such as vegetation, forestry, soil, water, and ice. These digital data when combined with the GPS result in a georeferenced image and elevation data derived from electromagnetic waves generated from the IFSAR active system. The products from this process is a digital radar orthoimage graphic and digital elevation radar data with a wide range of vertical accuracies from centimeters to tens of meters depending on sensors, reflights, processing techniques, and surface movement.

Airborne Laser Mapping is seeking a niche in the elevation market. Linear collection of elevation data using an airborne laser assembly combined with GPS and aircraft inertial reference system produce a three-dimensional positional XYZ digital file with horizontal and vertical components in latitude, longitude, and elevation which can be converted into a specific projection and rectangular coordinate system.

COMMONLY USED ALGORITHMS FOR ELEVATION DATA

1. BI-HARMONIC FILTERING - USED FOR INITIAL ESTIMATE OF SURFACE

FROM RAW DATA DERIVED BY SOME INTERPOLATION PROCEDURE. TENDANCY TO HONOR DATA POINTS IN MOST CASES.

A. RESULTS IN SURFACES OF MINIMUM CURVATURE AND TENSION. RESULTS IN VERY SMOOTH APPEARANCE ACROSS DATA AND EXHIBITS A CONTINUOUS TREND AWAY FROM THE DATA.

B. CONTOURS GENERALLY TEND TO HAVE GOOD CONNECTIVITY AND PARALLELISM.

2. <u>LAPLACEAN FILTERS</u> - USED FOR INITIAL ESTIMATE OF SURFACE FROM RAW DATA DERIVED BY SOME INTERPOLATION PROCEDURE.

A. RESULTS IN SURFACES WITH CONSIDERABLE CURVATURE PEAKING AT THE DATA POINTS.

B. CONTOURS GENERALLY TEND TO HAVE MORE CURVATURE.

3. <u>POLYGON OR PROXIMAL</u> - GENERATES GRID VALUES EQUAL TO THE NEAREST POINT VALUE.

A. RESULTS IN FLAT POLYGONAL AREAS SURROUNDING THE DATA POINTS.

B. A GOOD METHOD OF SURFACE DETERMINATION FOR MINING ACTIVITIES.

4. <u>DISTANCE WEIGHTING</u> - ASSIGNS MORE WEIGHT TO NEARBY POINTS THAN TO DISTANT POINTS.

A. DEPENDING ON THE SPATIAL LOCATION OF RAW DATA AMBIGUITY CAN RESULT DUE TO CHOICE OF WEIGHTING FUNCTION.

B. ARTIFACTS CAN OCCUR WHEN USED WITH AN UNEVEN DISTRIBUTION OF DATA POINTS. THIS CAN BE SEEN WHEN APPLIED IN AREAS OF FLAT TERRAIN WITH A SPARSE DISTRIBUTION OF DATA POINTS.

5. <u>SPLINES</u> - SOME POLYNOMIAL OF DEGREE M; LINEAR FOR M=1, QUDRATIC FOR M=2, CUBIC WHEN M=3 AND CONSIDERED BICUBIC WHEN USED IN THE THREE-DIMENSIONAL CASE.

A. SPLINE FUNCTIONS IN SPATIAL INTERPOLATION ARE FORMED BY PIECING TOGETHER SUCCESSIVE CURVE SEGMENTS INVOLVING RELATIVELY FEW POINTS AT A TIME AND SHOULD BE CLOSELY RELATED TO THE VALUE BEING INTERPOLATED; THEY ARE ANALYTIC; AND ARE FLEXIBLE. WHEN APPLIED TO A SPLINE SURFACE, SUCCESSIVE RECTANGULAR SURFACE PATCHES FORM A COMPOSITE SURFACE SIMILAR TO A PATCHWORK QUILT.

B. ADVANTAGEOUS FOR DENSE OR RECTANGULAR DATA, BUT CAN INTRODUCE ANOMALIES NOT IN THE ORIGINAL SURFACE DEPENDING ON THE INTERPOLATION AND BLENDING METHODS USED.

6. <u>UNIVERSAL KRIGING</u> - A STATISTICAL APPROACH TO A SURFACE OF IRREGULAR POINTS BY DETERMINING A SEMIVARIOGRAM, WHICH IS A FUNCTION RELATING THE COVARIANCE OF THE DIFFERENCE BETWEEN POINTS TO THE DISTANCES BETWEEN THE POINTS.

A. EFFECTIVE USE DEPENDS UPON THE PROPER SELECTION OF THE SLOPE OF THE SEMIVARIOGRAM, DEGREE OF POLYNOMIAL DRIFT, AND THE VARIANCE.

B. INTENSIVE COMPUTATION, BUT CAN PRODUCE AN ACCURATE GRID. A DIFFERENT SET OF EQUATIONS ARE USED FOR EACH POINT ESTIMATE IN DIFFERENT NEIGHBORHOODS.

7. <u>TRIANGULATED IRREGULAR NETWORK (TIN)</u> - THIS METHOD, IS NOT, AN INTERPOLATED REGULAR GRIDDED SURFACE, BUT RATHER A METHOD OF GENERATING A SURFACE FROM A SERIES OF X,Y,Z POINTS RESULTING IN A SERIES OF TOPOLOGICALLY STRUCTURED TRIANGLES WHICH CONNECT THOSE EXACT DATA POINTS TO THEIR NEIGHBORS.

A. A GOOD METHOD TO HONOR DATA COLLECTED ALONG RIDGES, DRAINS, AND SIGNIFICANT BREAKLINES OF CHANGING SLOPE ON THE TERRAIN SURFACE.

B. THE INITIAL SURFACE GENERATED WILL APPEAR PEAKED AT THE DATA POINTS AND HAVE ANGULAR SHARP CORNERS.

C. LOCATION OF THE DATA POINTS ON THE TERRAIN SHOULD BE ASSESSED BEFORE ANY APPLICATION OF SMOOTHING IS USED.

D. TOPOLOGY OF TRIANGLES ALLOWS FOR A VARIETY OF APPLICATIONS, SUCH AS SLOPE, SHADING, ASPECT, AND CONTOURS.

8. <u>CONVERGENT GRIDDING</u> - A PROCESS WHEREBY GRID NODE VALUES ARE CONVERGED-UPON THROUGH ONE OR MORE ITERATIONS OF SNAPPING CONTROL POINTS TO NEARBY NODES. THIS IS DONE USING A DISTANCE-WEIGHTING TECHNIQUE, SUCH THAT CONTROL POINTS CLOSER TO THE NODE HAVE A LARGER AFFECT ON THE OUTCOME OF THE NODE Z-VALUE. A BLENDING FUNCTION INSURES NON-COLLISIONS BETWEEN WEIGHTED Z- VALUES AND AVERAGES THE CONTROL POINT Z-VALUE USED.

A. A VERY GOOD ALGORITHM FOR HONORING HYPSOGRAPHIC DATA SUPPLEMENTED WITH HYDROGRAPHIC OR OTHER VECTOR DATA FOR FEATURE DEFINITION ENFORCEMENT, SUCH AS WATERLINE, DRAINS, OR GEOLOGIC FAULTS.

B. PROCESS IS VERY COMPUTATIONAL INTENSIVE DEPENDING ON CHOICE OF SIZE OF MOVING WINDOW AREA.

9. ITERATIVE FINITE DIFFERENCE INTERPOLATION - A DISCRETISED VERSION OF THE THIN PLATE SPLINE TECHNIQUE FOR WHICH THE ROUGHNESS PENALTY (DEFINED IN TERMS OF FIRST AND SECOND ORDER DERIVATIVES OF THE FITTED GRID) IS USUALLY A LINEAR CURVATURE OF THE FITTED SURFACE. THE ITERATION TECHNIQUE EMPLOYS A SIMPLE MULTI-GRID STRATEGY WHICH CALCULATES GRIDS AT SUCCESSIVELY FINER RESOLUTIONS, STARTING WITH A COARSE INITIAL GRID AND SUCCESSIVELY HALVING THE GRID SPACING UNTIL THE FINAL USER SPECIFIED GRID RESOLUTION IS OBTAINED.

A. A VERY GOOD ALGORITHM FOR HYDROLOGIC MODELING FROM CONTOUR DATA.

B. RIDGE AND DRAIN INFORMATION IS AUTOMATICALLY CALCULATED FROM THE CONTOUR DATA.

C. STREAM LINE DATA CAN BE USED WITH THE DRAINAGE ENFORCEMENT ALGORITHM TO INSURE A MORE ACCURATE PLACEMENT OF STREAMS.

Selected Bibliography

Abdel, S.M. 1982) Study of Factors Affecting Accuracy of Digital Terrain Models. *M.Sc. Thesis, Cairo University, Egypt*

Avecedo, W. (1991) First Assessment of U.S. Geological Survey 30- Minute DEMs: A Great Improvement over Existing 1-Degree Data. *Proceedings of the 1991 ACSM/ASPRS Annual Conference, Baltimore, Maryland*, 2:1-12.

Balce, A. (1986) Determination of Optimum Sampling Interval in Grid Digital Elevation Models Data Acquisition. *Proceedings ISPRS Commission III Symposium*, Finland, Int. Archieves of Photogrammetry and Remote Sensing, Vol. 26, Part 3.1, pp. 40- 55.

Braile, L. (1978) Comparison of four random to grid methods. *Computer & Geoscience*, Vol.4:341-349.

Douglas, D.H., and Peucker, T.K. (1973) Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. *Canadian Cartographer*, 10(2):112-122.

Garbrecht, J., and Starks, P. (1995) Note on the Use of USGS Level 1 7.5-minute DEM Coverages for Landscape Drainage Analyses. *Photogrammetric Engineering & Remote Sensing*, 61(5):519-522.

Gruen, A.W., and Baltsavias, E.P. (1987) High Precision Image Matching for Digital Terrain Model Generation. *Photogrammetria*, 42:97-112.

Hayes, J.G., and Halliday, J. (1974) The least squares fitting of cubic spline surfaces to general data sets. *Journal of the Institute of Mathematics and its Applications*, 14(1):89- 103.

Hunter, G.J, and Goodchild, M.F. (1995) Dealing with Error in Spatial Databases: A Simple Case Study. *Photogrammetric Engineering & Remote Sensing*, 61(5):529-537.

Hutchinson, M.F. (1991) The application of thin plate smoothing splines to continent-wide data assimilation. In: J.D. Jasper (ed), *Data Assimilation Systems*, BMRC Research Report No. 27, Melbourne: Bureau of Meteorology, pp.104-113.

Jenson, S.K. (1991) Applications of Hydrologic Information Automatically Extracted from Digital Elevation Models. *Hydrological Processes*, 5:31-44.

Jenson, S.K., and Domingue, J.O. (1988) Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. *Photogrammetric Engineering & Remote Sensing*, 54(11):1593-1600.

Lam, N.S. (1983) Spatial interpolation methods: a review. *American Cartographer*, 10(2):129-149.

Lue, Y., and Novak, K. (1991) Simultaneous Extraction of Digital Elevation Models and Orthophotos. *Technical Papers, ACSM- ASPRS Annual Convention*, Vol.5:412-418.

Mirante, A., and Weingarten, N. (1982) The radial sweep algorithm for constructing triangulated irregular networks. *IEEE Computer Graphics and Applications*, 2(3):11-21.

Norvelle, F.R. (1992) Window shaping and DEM corrections. *Photogrammetric Engineering and Remote Sensing*, 58(1):111- 115.

O'Callaghn, J.F., and Mark, D.M. (1984) The Extraction of Drainage Networks form Digital Elevation Data. *Computer Vision, Graphics, and Image Processing*, 28:323-344.

Schurt, G.M. (1976) Review of Interpolation Methods for Digital Terrain Models. *The Canadian Surveyor*, 30(5):389- 411.

Sibson, R. (1981) A brief description of natural neighbour interpolation. *Interpreting Multivariate Data*, (Editor V. Barnett). John Wiley and Sons, Chichester, pp. 21-36.

U.S. Geological Survey (1990) *Digital Elevation Models Data Users Guide*, Reston, Virginia, 51 p.

Yoeli, P. (1977) Computer executed interpolation of contours into arrays of randomly distributed height points. *Cartographic Journal*, 14(2):103-108.

Yoeli, P. (1986) Computer executed production of a regular grid of height points form digital contours. *American Cartographer*, 13(3):219-229.

Application of CAD Framework Techniques to Systems Integration in Environmental Modeling

Sally Kleinfeldt (1), Jonathan Deckmyn (1), Pieter van der Wolf (2), Olav ten Bosch (2), and Claudio Paniconi (1)

1. ***CRS4, Via Nazario Sauro 10, 09123 Cagliari, Italy***
2. ***Delft University of Technology, Faculty of Electrical Engineering / DIMES, The Netherlands***

Abstract

Systems integration is a growing concern for environmental researchers faced with an increasing volume of observed and simulated data and with increasingly sophisticated models and tools such as geographic information systems and scientific visualization techniques. These users need targetted, application-specific systems incorporating selected functionality from models, GIS, visualization, and analysis systems, and they need to operate on data of diverse types and formats --- but they do not generally need or wish to become experts on the intricacies of each of these tools or components. A robust integration methodology is required that will enable a user to quickly build a customized system from various available components, many of which are proprietary and cannot be altered.

Environmental researchers have made several attempts to integrate desired components within (often proprietary) GIS. The difficulties of this approach --- particularly for applications involving 3 or 4 dimensions --- have been well-documented. Another approach has been to create in-house GIS which are more amenable to model integration requirements. This alternative, however, is not realistic for the average environmental research institute or department.

The purpose of this paper is to draw attention to the substantial literature that already exists on an analogous problem: that of integrating uncooperative, often proprietary tools in the world of computer aided design (CAD). In CAD, software infrastructures for building integrated environments, called CAD frameworks, have been developed. It is our feeling that environmental researchers could profitably adopt techniques that have been developed in the CAD field.

CAD framework techniques have been used extensively as the software ``glue'' that can turn a collection of individual tools into an effective and user-friendly integrated engineering environment. CAD frameworks provide a wealth of functions that not only support the construction of integrated environments, but also assist users in operating such environments

[*van der Wolf, 1994*].

Historically, the first role allotted to CAD frameworks was that of common data repository, or *design database*. Data which is common to a number of CAD tools is stored only once in the repository, from where it can be used as input for all tools. This promotes *tool interoperability*. We emphasize that a design database merely provides a facility for storing and retrieving data; actual tool interoperability would require common formats for the data in order to avoid translation steps.

Subsequently, more functions were added to support the management of design data. CAD frameworks started to fulfill a second role, that of *design data manager*. A design data management system maintains information about the structure and status of the design to provide management support and enforce constraints on the design process. Graphical browse facilities were introduced to visually present this information to the user. In the CAD framework community this information about the design data is called meta design data, or simply *metadata*.

The third major role allotted to CAD frameworks was that of *design process manager*. With the increasing number of tools in today's CAD systems, there is a growing need to support the user in correctly executing these tools to perform tasks. On top of such a tool management service, the framework may provide *design flow management*, which assists the user in correctly and efficiently performing design activities according to a locally defined design procedure.

As a first example of applying CAD framework concepts to an integration problem in the environmental sciences, we will describe a system to provide data management and browsing services for a collaborative hydrologic research effort. We needed to create an environment within which all project data could be managed, browsed, and retrieved for further study. More importantly, we wanted the browsing feature to allow in depth examination of the available data using various visualization techniques, some from GIS tools and some from scientific visualization systems.

To address our data management needs, we created an independent, centralized repository. To address tool interoperability and reduce the complexity associated with the heterogeneous nature of the data, the repository data is stored in standard formats (Freeform and HDF). To address our browsing needs, we created a graphical user interface (HYBROW, for Hydrologic Browser) [*Kleinfeldt et al., 1996*], shown in Figure 1, which supports visualization of the contents of the repository. We have encapsulated functionality from tools such as GRASS, AVS, Iris Explorer, Gnuplot, and ftp. HYBROW was implemented in C and Tcl/Tk, a simple, public domain scripting language with built-in facilities for creating X Windows ``widgets''.

A second example of applying CAD framework concepts is based on NELSIS, a design flow based CAD framework developed at the Delft University of Technology [*ten Bosch, 1995*]. The idea behind design flow management is to transform the informal idea of the structure of the design process as it resides in the user's mind into a formal description (a design flow) which is then used to further assist the user. In our case the design flow describes the basic steps in the modeling process, such as preparation and checking of inputs, model execution, and interpretation and analysis of outputs. The framework allows complex queries in an intuitive way. It doesn't impose any restrictions on data organization, nor on the use of graphical user interface tools.

Our first results using the NELSIS framework are described in Figure 2. The elaboration of a complete ``NELSIS-flow'' for a complex hydrologic model is in progress.

References

Kleinfeldt, S., J. Deckmyn, B. Cosyn and C. Paniconi, GIS and scientific visualization for hydrologic simulation. To appear in: *Proceedings of HydroGIS '96, International Conference on Application of Geographic Information Systems in Hydrology and Water Resources Management, Vienna, April 16-19, 1996.*

ten Bosch, O., *Design Flow Management in CAD Frameworks*. CIP-DATA Koninklijke Bibliotheek, Den Haag, 1995.

van der Wolf, P., *CAD Frameworks: Principles and Architecture*. Kluwer Academic Publishers, Boston/Dordrecht/London, 1994.

- Figure1: HYBROW
- Figure2: NELSIS

Jonathan Deckmyn
e-mail: jdeckmyn@crs4.it

Kenneth L. Driese and William A. Reiners

A preliminary GIS-based biogeographic exploration of the relationship between vegetation distribution and environmental variables in Wyoming, USA

A digital map depicting the distribution of land cover in Wyoming has been completed as part of the National Biological Service (NBS) Gap Analysis Program (Gap). This map, along with a suite of spatially-explicit climate and soil variables, was used to explore vegetation-environmental relationships, and the effect of using environmental data at different resolutions to predict the geographic distribution of vegetation. Specifically, we used mean monthly temperatures from January and July (10.9 km resolution), annual precipitation amount (4.7 km), a winter precipitation index (4.7 km), growing degree days (10.9 km), potential evapotranspiration (PET) (10.9 km), soil salinity (100 m) and soil surface texture (100 m) to predict the distribution of 26 Gap cover types (e.g., lodgepole pine forest) and of 14 aggregated formation level types (e.g., evergreen needleleaf forest with rounded crowns). Precipitation and temperature data were derived from PRISM (Daly 1994), growing degree days and PET from Neilson (1995), and soil data from STATSGO (Soil Conservation Service 1991).

Statistical environmental signatures for each type, created using routines in the GRID module of Arc/Info (ESRI, Redlands, CA), were calculated for vegetation at full resolution (100 m) using environmental variables at each of the 3 resolutions of the source data (10.9 km, 4.7 km and 100 m). The statistical signatures were used to produce maximum-likelihood classifications of land cover for Wyoming. Overall classification accuracy was calculated from error matrices that tabulated correspondence between predicted and mapped land cover. Accuracy was calculated by dividing the number of correctly classified cells by the total number of cells in the domain. This method did not consider errors of commission and omission but was sufficient for comparisons between tests.

Overall classification accuracy for the predicted distribution of the 26 Gap cover types was 13% using environmental data at all 3 resolutions. The classification accuracy for the prediction of aggregated formation level types was 25% with environmental data at 10.9 km resolution, and slightly higher (30%) for prediction with 4.7 and 100 m data. These results suggest th at our ability to predict land cover at both the cover type and the formation level with only these environmental variables is weak, and that other environmental data (topography, soil type, etc.), or models that include processes (i.e., water balance), are needed to improve our predictive power. The results were not sensitive to the resolution of the environmental data, indicating that the coarsest environmental data used in the analyses (January and July mean temperature, PET, and Growing Degree Days) overpowered the influence of the finer resolution precipitation and soil derived data.

References

Daly, C, R.P. Neilson and D.L. Phillips. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33:140-158.

Nielson, R.P. 1995. A model for predicting continental-scale vegetation distribution and water balance. *Ecological Applications* 5(2):362-385.

Soil Conservation Service. 1991. State Geographic Data Base (STATSGO). Data Users Guide. U.S.D.A. Soil Conservation Service Misc. Publ. No. 1492. 88 p. plus digital data.

Kenneth L. Driese
Department of Botany
University of Wyoming
Laramie, WY 82071-3165
phone: 307-766-2484
fax: 307-766-2851
e-mail: kdriese@uwyo.edu

William A. Reiners
Department of Botany
University of Wyoming
Laramie, WY 82071-3165
phone: 307-766-2235
fax: 307-766-2851
e-mail: reiners@uwyo.edu

Dynamic Linkages Between GIS And A Physically Based, Spatially Distributed Hydro-Geomorphic Model CLAWS

Jinfan Duan[1], Gordon E. Grant[2] and Chaur-Fong Chen[3]

[1]*Department of Forest Engineering, Oregon State University, Corvallis, OR 97331*
[2]*Pacific Northwest Research Station, U.S. Forest Service, Corvallis, OR 97331*
[3]*Department of Bioresource Engineering, Oregon State University, Corvallis, OR 97331*

The last decade has seen an increasing emphasis on the need to predict spatially varying hydrologic and geomorphic responses at quite fine resolutions using spatially distributed models. The CLAWS, a physically based spatially distributed hydrologic and geomorphic 4D model, is designed to simulate the hydrologic and geomorphic responses of forest watersheds to landscape and climate changes. This model integrates the major hydrologic and geomorphic processes such as snow accumulation and melting, interception, evapotrasporation, subsurface flow and overlandflow, hillslope stability and sediment transport with vegetation management options. Time and expertise required for data preparation and model parameter estimations, and post-run spatial analysis and data visualization can be prohibitive. A Geographic Information System has been dynamically linked with CLAWS to address the watershed response to varying soil, vegetation and particular topographic attributes. This expedites both the pre-run terrain analysis based on Digital Elevation Model data and other spatial layers of soil and vegetation characteristics, and the post-run spatial data visualization which include elemental flow, soil moisture distribution, factor of safety, debris flow path and snow distribution output at any point during the simulation based on GIS readable files and accompanying AMLs.

Ralph Dubayah

The Tyranny of Scale and the Multifractal Paradigm

One of the major uses of GIS is to provide an environment that facilitates distributed environmental modeling. The distributed input fields required to drive such modeling efforts increasingly are being derived from remotely sensed data. The integration of these data within a GIS environment can be difficult, primarilly because of issues of scale. Remotely sensed and model output fields share an important property concerning their spatial variability: both are inherently scale limited. For remotely sensed observations, little information can be inferred below the resolution of the sensor. For distributed outputs, the model grid spacing determines the smallest spatial scale at which fields may be realized. The limitations imposed by resolution and grid spacing, or their interaction, can seriously hamper modeling efforts over large areas, especially when the process being modeled is spatially autocorrelated. Such processes often scale non-linearly such that the moments of the field obtained at one spatial scale may be significantly different from that obtained at a larger or smaller scale. The common mode of exploration has been to realize fields at fine scales, aggregate, and compare the results with those realized at coarser scales. The hope is that either linear or otherwise parameterizable relationships may be found that link field statistics, at least for some limited range of scales. Otherwise, if the fields do not scale linearly, or if we do not know how to model the non-linearity, we are forced to model at the scale of interest, regardless of its practicality -- this is the tyranny of scale.

It has been shown recently that non-linear spatial variability among scales may be a consequence of fields that are generated by cascade processes, where the total flux is the result of some cascading down of large scale fluxes to successively smaller and smaller scales. Cascade processes can produce fields which exhibit "multiscaling" or "multifractal" behavior, characterized by temporal or spatial scaling exponents which are non-linear with respect to the order of statistical moment. Such scaling exponents, when they exist, may be used to model a process at one scale and infer its variability at any other scale. As such, the multifractal framework is particularly well suited to the issue of scaling within the context of remote sensing and GIS-based modeling. In this poster I outline its theoretical basis and an accompanying data analysis methods, as well as illustrating its relevant application to distributed modeling and remotely sensed fields such as microwave soil moisture and surface temperature. Although no panacea, the multiscaling paradigm at least suggests the possibility of freeing the modeling process from the constraints imposed by the observational scale of inquiry.

Ralph Dubayah
Geography Department
University of Maryland, College Park MD 20742
email: rdubayah@geog.umd.edu
Telephone: (301) 405-4069

Fax: (301) 314-9299

Thomas C. Edwards, Jr. and Scott D. Bassett

A Spatial Analysis of Protection Afforded Land-Cover and Terrestrial Vertebrate Diversity in Utah

---

Abstract:

We evaluated the extent to which newly proposed wilderness for Utah, in conjunction with other Federal, State and private lands, affords protection to Utah's land-cover types and terrestrial vertebrate diversity. All analyses used the Utah Gap Analysis Environmental Information System (Edwards et al. 1995), a USDI National Biological Service program designed to develop spatially explicit environmental information systems for use in resource planning. Our emphasis on cover-types was simple representation. For terrestrial vertebrates we considered 3 levels of analysis: (1) representation in space; (2) redundancy in coverage; and (3) area, evaluated in terms of separate parcels of reserved land and proposed wilderness having sufficient area for 1 to 1,000 species-specific home ranges.

Six major cover-types were evaluated, including forest, woodlands, shrublands, herbaceous communities, barren lands (<5% vegetation cover), and aquatic communities. Proposed wilderness in conjunction with existing reserves (e.g., national parks, current wilderness, refuges) covers 3.6% to 14.5% of the statewide total of each class. Only two classes, barren lands and woodlands, exceed or come close to a minimum 12% in reserve status suggested by the International Union of Conservation of Nature and Natural Resources (IUCN) (c.f. Hummel 1989).

Opportunity exists for Utah to increase its protection of land-cover diversity by increasing the amount of area in the proposed wilderness bill for the 4 cover-types not meeting the IUCN threshold. Unfortunately, no adequate scientific evidence exists expressing the area necessary to protect cover-types, and the few estimates available range from 5% to as high as 99% (Ryti 1992). Given this variability in area requirements, it is difficult to evaluate the protection afforded Utah's mapped cover-types. Using the IUCN estimate of 12% since it refers to biomes and can be considered analogous to the cover-types evaluated, only barren lands exceeds the 12% threshold. Woodlands, with an estimated 11% in reserve status, is close enough to considered, for all intent and purpose, adequately covered. The remaining four classes are roughly 6% to 9% short of the 12% threshold and consideration of additional area in any Utah wilderness proposal for these four classes is warranted.

Spatial coverage merely determines if any piece of a species' critical or high value habitat is contained in reserved land. By itself, the proposed wilderness area does not include any predicted critical or high value habitat for 13% of the birds, 24% of the mammals, 18% of the reptiles, and 29% of the amphibians in the state. When coupled with existing reserved land in the state, however, only 15 of Utah's 525 (3%) terrestrial vertebrates do not have some

portion of the predicted distribution of their critical and high value habitat included.

Redundancy in coverage is a measure of protection from localized catastrophe. As redundancy in coverage increases from 1 to 5 separate parcels of land, the percent of species not covered increases for all taxa. Redundancy in coverage was poorest for proposed wilderness, with >35% of all species not having coverage in 3 or more parcels of land. Reserved land fared better, with <10% of amphibians, birds and mammals not having coverage in 3 or more parcels. Combining reserved land with proposed wilderness reduces the percent of species not covered slightly, but still leaves about 25% of reptiles not having redundancy in coverage of 3 or more parcels of land.

Area, as measured by species-specific home range requirements, provides a general estimate of the ability of parcels of land to maintain animal populations. The percent species by taxon not having 100 or more home ranges in proposed wilderness ranges from a low of 40% in birds to 85% in amphibians. Current reserved land lacks sufficient area for 100 or more home ranges for about 20% of birds and mammals, 35% of reptiles and >65% for amphibians. Values are similar when current reserved land is combined with proposed wilderness.

Different interpretations of the amount of Utah's terrestrial vertebrate diversity not protected occur depending on the method used to evaluate protection. Simple Gap Analysis, the first of the three methods presented here, relies heavily on representation in space (Scott et al. 1994). Consequently, it is possible to consider "protected" a species for which sufficient area for a single home range is not present. Under these circumstances Utah's current reserved land plus proposed wilderness provide protection for about 97% of the state's terrestrial vertebrates. More ecologically complex models change the percent of species considered protected. Redundancy in coverage, a form of protection from localized catastrophe (Noss and Cooperrider 1994), reduces the percent of species protected to about 75%, with reptiles faring the worst. When species-specific home range requirements are built into the models, <50% of Utah's terrestrial vertebrates can be considered protected on reserved land.

The approaches presented here represent a first attempt to describe protection afforded land-cover and terrestrial vertebrate diversity in Utah. Once an assessment of current protection is accomplished, algorithms for optimizing the selection of additional reserved land to incorporate unprotected components can be applied (see Csuti et al., in press). Although all these algorithms emphasize biodiversity in different fashion, they all ignore practical constraints like social perceptions, land ownership and economic conflicts. To be successful, future approaches to the spatial placement of reserved land must incorporate social as well as biological factors.

References:

Csuti, B., Polasky, S., Camm, J.D., Downs, B., Hamilton, R., Huso, M., Kershaw, M., Keister, A.R., Pressey, R.L., Sahr, K., and Williams, P.H. (In press) A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon. *Biological Conservation*.

Edwards, T.C., Jr., Homer, C.G., Bassett, S.D., Falconer, A., Ramsey, R.D., and Wight, D.W. (1995) *Utah Gap Analysis: an environmental information system.* Technical Report 95-1, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, Utah.

1189pp + 2 CD-ROMs.

Hummel, M, ed. (1989) *Endangered spaces: the future for Canada's wilderness*. Toronto: Key Porter Books.

Noss, R.F., and Cooperrider, A.Y. (1994) *Saving nature's legacy: protecting and restoring biodiversity*. Washington, D.C: Island Press.

Ryti, R.T. (1992) Effect of the focal taxon on the selection of nature reserves. *Ecological Applications* 2:404-410.

Scott, J. M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Caicco, S., Groves, C., Edwards, T.C., Jr., Ulliman, J., Anderson, H., D'Erchia, F., and Wright, R.G. (1993) Gap analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs No*. 123.

Authors:

Thomas C. Edwards, Jr., National Biological Service, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, UT 84322-5210, 801-797-2529 (PH), 801-797-4025 (FAX), tce@nr.usu.edu (E-MAIL)

Scott D. Bassett, Department of Fisheries and Wildlife, Utah State University, Logan, UT 84322-5210, 801-797-3892 (PH), 801-797-4025 (FAX), sbassett@nr.usu.edu (E-MAIL)

A GIS decision support system for fire and alien weed management in the nature reserves of the Western Cape Province, South Africa: spatial simulation to application

Dean H.K. Fairbanks, B.W. van Wilgen, D.H. McKelly, K. Feast, and D.H. Le Maitre
CSIR, Division of Water, Environment and Forestry
P.O. Box 395
Pretoria 0001
South Africa
Tel: +27 12 841-2489
Fax: +27 12 841-2689
Internet: dfairban@csir.co.za

Conservation authorities, in the Western Cape Province of South Africa, are dependent on relevant information to make ecologically sound decisions for their nature reserves. Twenty years of Fynbos ecology research and recent spatial spread models have shown that the monitoring of alien weed encroachment and fire management are the most critical information factors needed for sound conservation and water impact management. Until recently the data and management strategies have been conducted in a quasi-GIS based approach. New gains in approaching both graphical and modeling scenarios on PC-based spatial decision support systems have lead to hands on management and monitoring support for the field rangers and office managers. This poster presents an initial GIS spatial spread modelling approach to understanding the effects of the lack of management on alien weed spread within catchement water resources and the subsequent PC-based support system (Conservation Management System) created to manage and monitor the reserves and catchment resources. Simple rule-based and real-time dynamic models are used for presentation and decision making. Current applications are: prioritization of areas for burning, monitoring the success of fire management, mapping fire hazard for fire control planning, management strategies for controlling alien weed spread, and the production of management summaries and statistics. Support, management, and subsequent monitoring of reserves can be achieved in an effective manner by providing an integrative GIS decision support system.

Natural resource planning by way of a geographic spreadsheet modelling approach

Dean H.K. Fairbanks and Mike P. Adam
CSIR, Division of Water, Environment and Forestry
P.O. Box 395
Pretoria 0001
South Africa
Tel: +27 12 841-2489
Fax: +27 12 841-2689
internet: dfairban@csir.co.za

Many models created for managing natural resources are based on a spreadsheet approach. The spreadsheet based mathematical models used by planners for resource management often tend to leave out the spatial component of the problem. The current spreadsheet models are aspatial in a sense that they are operated on spreadsheets from attributed spatial data. Unfortunately the management solutions given do not always allow for a spatial management solution, or if they do then query and display of these solutions is not incorporated into the answer. This is typically seen when planning landscapes for development vs. conservation, managing biological populations, water resources, etc. In South Africa there has been a drive to understand how to link spreadsheet based mathematical natural resource analysis solutions to a GIS and how this can be constructed into PC-based decision support tools. This poster presents the development transition from a basic spreadsheet approach to a geographic spreadsheet approach to modelling natural resources. Examples are shown from two management disciplines: land use planning and water resources. The Land Use Planning System is based on the principal that one can change land cover/land use types to other set categories and that these other land uses in turn have an impact both socio-economically and environmentally in any planning decisions. The user is able to change parcels on the map which then allows a simple spreadsheet model to calculate the effect of that change from this planning maps and statistics maybe collected in an efficient manner. The Afforestation Runoff Impact Modelling System works on the same principal that by selecting catchments and adding or taking away a variety of forestry species by specifying total area, site quality, rotation length, and rainfall zone, this will have an effect on the runoff resources of that catchment and its neighbours through an empirical reduction model. Simple spreadsheet models are used to calculate these changes and show them graphically on the map and collect statistics. The current PC-based GIS systems allow for powerful and flexible decision support systems to be created. This in turn has allowed traditional natural resource models done on spreadsheets to be transformed into geographic spreadsheet planning tools.

The Response of Vegetation to Change of Annual Rainfall in the Sahel Region of Africa, and its Dependence on Soil Type.

George W. Fisher (Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore MD, 21218) and Elissa Levine (Goddard Space Flight Center, Code 923, Greenbelt, MD 20771).

Modeling the response of arid and semi-arid lands to changes in rainfall is essential for predicting the consequences of global change in dry regions of the world.

Tucker et al (1991) used changes in normalized difference vegetation index (NDVI) for the Sahel region of Africa to show that the southern boundary of the Sahara moves southward in years of low rainfall and retreats northward in years of higher rainfall. Using the IDRISI GIS system, we have divided the Sahel into subareas characterized by different soil types (based on a digital version of the FAO soil map of the world), produced separate NDVI maps for each soil type, and prepared histograms of year-to-year differences in NDVI for each subarea. Our results show that variations in rainfall produce different changes in NDVI on different soil types; in general, sandy soils show a larger change in NDVI for a given change in rainfall than do soils with higher clay content. It appears that clayey soils retain more moisture in dry years, and gain less moisture in wet years, than do sandy soils, and that these differences affect the abundance of vegetation.

Consequently, attempts to predict the effects of climate change in arid and semi-arid regions must take into account differences in soil type and the effects of those differences on the response of vegetation to changes in rainfall.

Jeremy S. Fried, Mark O. Zweifler, Michael A. Gold and Dan G. Brown

GIS approaches to targeted siting of riparian buffer strips: trade-offs between realism and complexity

---

Abstract

Vegetated riparian buffer strips (RBS) represent a promising approach to impeding the delivery of non-point source (NPS) pollutants into streams, thereby protecting or enhancing water quality. Watershed managers and planners are interested in deploying RBS in the field, but rarely have the resources to establish them along every meter of stream, creek and drain. At MSU, a statewide issues research project is addressing the issue of targeted siting of RBS via GIS modeling, a survey of riparian landowners, and interviews with key institutional players.

The ideal GIS model would prioritize locations for RBS installation by both type and magnitude of potential non-point source pollution problem and likelihood of RBS effectiveness, operate at a high enough resolution to guide decisions about efficient action on individual parcels, and be parsimonious in its requirements for data and skilled analysts. Although others (e.g., Inamdar et al., 1993) have sought to prioritize sites for edge-of-field buffers using GIS representations of slope and land cover, GIS modeling has not yet been integrated into RBS siting analysis. This poster summarizes several GIS approaches to riparian buffer strip siting (dubbed Models I-V) ordered along a continuum of complexity and information richness (of both inputs and outputs). The first three have been constructed for a sub-watershed of Sycamore Creek, near Lansing Michigan; the others are described as work in progress.

Common to all models is an investigative riparian buffer. This conceptualization greatly reduces the areal extent of the NPS management problem by restricting attention to that portion of the watershed through which NPS pollution gains entry to streams and in which most NPS pollution problems are generated.

Model I, an overlay of a fixed width buffer on aerial photo-interpreted land use, should be popular with those who consider land use to be the primary consideration in identifying water pollution sources. Fixed-width delineation of the riparian zone (e.g. 250 meters on each side of the creek) has little basis in hydrology, but is easy to construct. Land uses within the riparian buffer can then be ranked by probable contribution to water pollution and/or degree of inherent pollution mitigation capacity. For example, row crops might be rated high based on probable contribution; forest might be rated low by virtue of filtration capacity.

Model II, a multiplicative, map algebra model (Sivertun et al., 1988), uses ratings based on

distance to stream, slope, soil K factor, and land use to calculate a critical value index for each cell. While its suitability index approach may be familiar to potential users, the model focuses on source areas. The model's tendency to rate isolated locations which are quite far from the stream as critical demonstrates that even with the inclusion of distance to stream, NPS pollution delivery is not well represented. Sophistication and accuracy may be somewhat greater than for Model I, as are its data requirements.

Based on the hypothesis that both detachment and transport are required for pollution to enter a waterway, Model III relies on the DYNWETG component of the Terrain Analysis for the Environmental Sciences (TAPES) software (Gallant and Wilson, 1996) to calculate dynamic wetness index ($DWT=\ln(A_e/S)$), with or without distributed soils parameters, and stream power ($PWR=A_eS$), (A_e = effective upslope contributing area, S=slope), as indices of detachment and transport. Fuzzy membership functions were defined for each index, such that the function is 1 for the 0-50th percentile, decreases linearly to 0.01 at the 95th percentile, and is 0.01 above that point. A unit cost surface was generated by calculating the product of these fuzzy sets (i.e., the fuzzy intersection). The cost surface values range from 0.0001 to 1 and are least where there is both high dynamic wetness and high stream power. Distance from the stream can then be accumulated over this cost surface to generate a cost distance map. The lowest *n* percent of values in the cost distance map can then be reclassified to generate a binary investigative riparian buffer, with the selection of *n* dependent on situation specific factors. Land use coincident with this buffer can serve as the basis for further prioritization of areas within this buffer based on NPS pollution potential, e.g., the binary investigative buffer can be used to clip a land use layer to create a hydrologically defensible, variable width, land use cognizant, "critical" buffer spanning the stream (Figure 1).

Model IV, not yet constructed, would essentially be a refinement of model III which would integrate DYNWETG's dynamic wetness algorithm into the TAPESG software to take advantage of TAPESG's Digital Elevation Model's (DEMON) flow routing option. Comparisons of output from models with D8 and DEMON flow routing algorithms conducted by the authors and others (e.g., Costa-Cabral and Burges, 1994) confirm that indices based on DEMON produce far more realistic looking maps of hydrologic flow, and eliminate the artifacts of "flow shadows" produced by D8.

With some additional programming, Model IV could be extended to incorporate a distributed coverage of weights representing pollution potential derived from a land use coverage into the flow accumulation calculations. Dubbed Model V, this representation would assign greater weight to surface flow with high stream power originating from catchments with higher concentrations of NPS generating land use.

Two types of validation were attempted: 1) a GPS based stream bank survey for concentrated runoff and ponding and 2) manual interpretation of USGS topographic quadrangle maps (Figure 1). Ultimately, the conveyance of pollutants across the riparian corridor and into the stream must be monitored during storm run-off events to assess the reliability of any of these models.

While technically deficient, Model I may still be a good choice as a first step towards an investigative buffer, and can be generated with minimal GIS expertise; however, up-to-date land use may be difficult to obtain in many areas. Model II is more sensitive to soils and terrain, but requires additional GIS coverages and somewhat greater analytic acumen. Either could be constructed by a typical GIS consulting firm. The far greater complexity of Models III-V, and the greater difficulty inherent in mastering the TAPES software make these more appropriate choices for managers with access to sophisticated GIS analysis support. Additional monitoring is needed to compare the efficacy of these models.

Implementation of any of these models poses real challenges for watershed managers, a clientele likely to have little or no experience using, building or fitting GIS models. After validation and testing of these models, our ultimate objective is to compare the predictive power of the high cost/accuracy/resolution models such as III-V with that of more affordable and parsimonious methods that would be accessible to watershed managers. Managers could then select the approach best suited to their information needs, analytic capabilities, and budget.

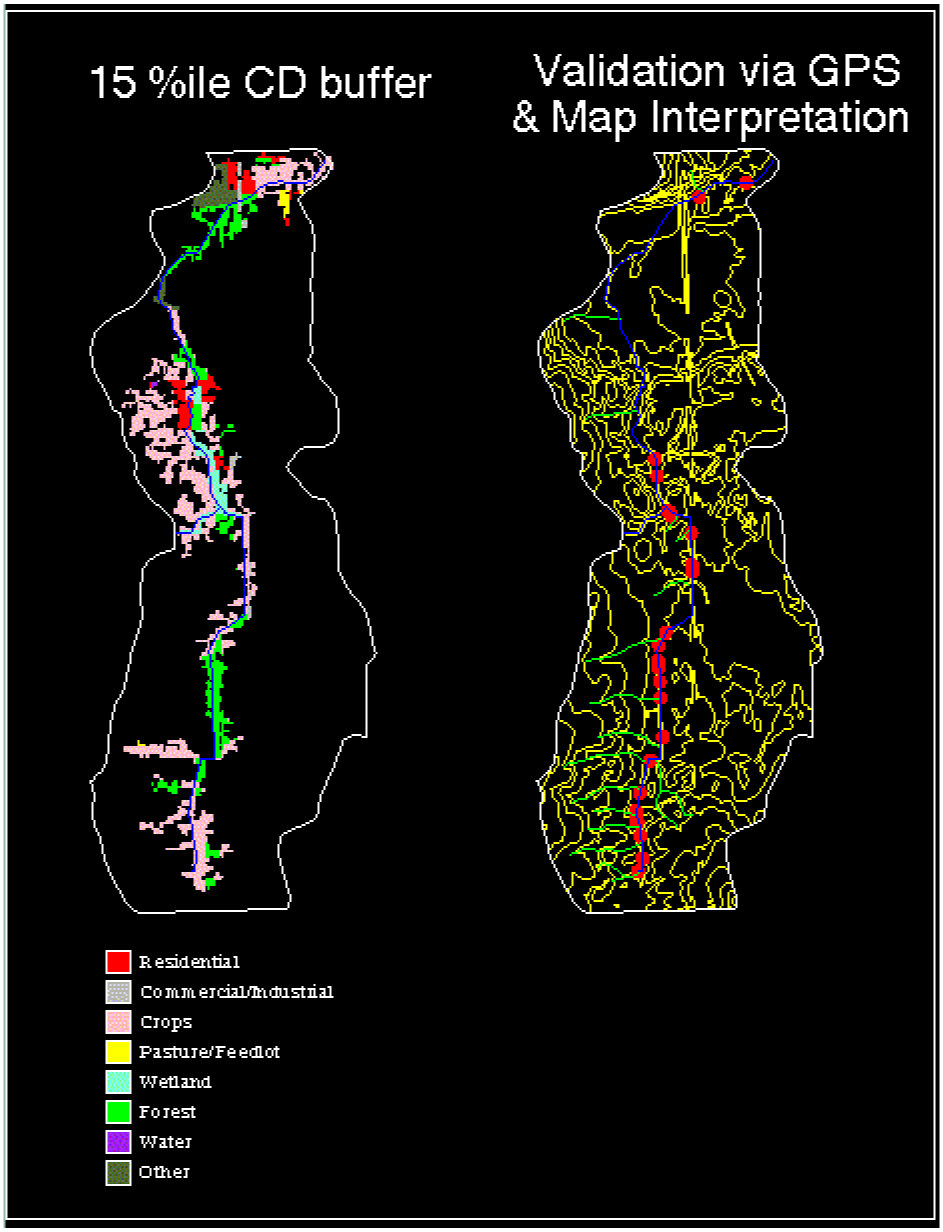


Figure 1. Left: Investigative buffer generated using Model III and a 15th percentile cost distance threshold for Barnard Drain, superimposed on land use. Right: Potential pollution entry sites located by the stream bank survey are represented as points; problem (concentrated flow) areas as determined from interpretation of topographic maps are represented by flow lines running the length of the presumed contributing areas.

Literature Cited

Costa-Cabral, M.C. and S.J. Burges. 1994. Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. Water Resources Research 30:1681-1692.

Gallant, J.C. and J.P. Wilson. 1996. TAPES-G: A grid-based terrain analysis program for the environmental sciences. Computers and Geosciences 17(3):413-422.

Inamdar, S.P., S. Zacharias, C.D. Heatwole, and T.A. Dillaha. 1993. Spatial placement of filter strips using a GIS. Presented at the December, 1993 meeting of the American Society of Agricultural Engineers, Paper No. 93-3560.

Sivertun, Reinelt, and Castensson. 1988. A GIS method to aid in non-point source critical area analysis. Int. J. of Geographical Information Systems 2:365-378.

Author Information

Jeremy S. Fried
Michigan State University
Department of Forestry
126 Natural Resources
East Lansing, MI 48824
517-432-3352
jeremy@msu.edu

Mark O. Zweifler
Michigan State University
Department of Forestry
126 Natural Resources
East Lansing, MI 48824
517-432-3537
mark@okemos.for.msu.edu

Michael A. Gold
Michigan State University
Department of Forestry
126 Natural Resources
East Lansing, MI 48824
517-353-4751
mgold@msu.edu

Daniel G. Brown
Michigan State University
Department of Geography
412 Natural Science
East Lansing, MI 48824
517-353-9811
brownda@pilot.msu.edu

Using Remote Sensing Analysis of Landsat Data to Evalutae an Integrated Socio-Economic Model of Deforestation in the Amazon

Robert C. Frohn, Kenneth C. McGwire, John E. Estes, Virginia H. Dale

Abstract:

The effectiveness of an integrated socioeconomic and ecological simulation model for estimating patterns and rates of deforestation in Rondonia, Brazil is evaluated using remote sensing analysis of Landsat data and Geograhic Information Systems. This study evaluates the model's ability to estimate spatial and temporal patterns of deforestation by comparing landscape metrics between image classifications and model outputs. The landscape metrics used include Percent Cleared, Contagion, and Fractal Dimension. Results indicate that the rates and spatial patterns of deforestation are similar between the model outputs and the Landsat data. Differences in clearing patterns between the model and Landsat data are due in part to localized farming obstacles, topography, and patchiness of clearings within lots which are not represented in the model. This project demonstrates the integration of remote sensing and geographic information systems for the modeling and analysis of the processes involved in tropical deforestation.

Yan Zhou, Chris Fulcher, Tony Prato

Watershed Management Decision Support System

There is a growing consensus that an effective way to control nonpoint source pollution and enhance the long-term sustainability of agriculture and rural communities is through locally-based planning and management at the watershed scale. Coordinated resource management of a watershed requires the simultaneous consideration of physical and socioeconomic interrelationships and impacts. In order to address these considerations, it is necessary to integrate a large amount of spatial information and knowledge from several disciplines. To be useful, the information and knowledge must be made available to decision makers in a rational framework.

Advances in remote sensing, geographic information systems (GIS), multiple objective decision making, and physical simulation make it possible to develop user-friendly, interactive, decision support systems for watershed planning and management. The goal of the study is to incorporate these advances by designing a user-friendly, interactive watershed management decision support system (WAMADSS) that identifies the relative contribution of sub-watershed areas to agricultural nonpoint source pollution and evaluates the effects of alternative land use/management activities and practices (LUMAPs) on farm income, soil erosion and surface water quality at the watershed scale. LUMAPs to be included in WAMADSS are: crop rotations, tillage practices, conservation practices (grass waterways, terraces), pollution prevention practices (timing, rate and method of application of fertilizers and pesticides) and other landscape elements such as improved vegetative cover in riparian areas. The decision support system (DSS) adopts a landscape perspective which is a way to view interactive parts of a watershed rather than focusing on isolated components.

The watershed management decision support system has three major components: a GIS, a modeling system, and a graphical user interface (GUI). ARC Macro Language (AML) is used to construct the GUIs which interface the simulation models and the economic model in a seamless decision support system framework. AML handles all simulation-related activities, including generating input files, executing the environmental models, and viewing results in the GIS.

Geographic Information System: ARC/INFO software is used in WAMADSS to significantly improve the user's ability to manipulate the spatial and non- spatial data needed to evaluate alternative watershed management plans. This approach enhances the "best judgment" decisions offered by conventional environmental models such as AGNPS, SWAT and CARE. ARC/INFO contains modules for interfacing models in a decision support system. Specifically, the spatial analysis modules in ARC/INFO generate the needed physical parameters such as slope, aspect, and slope length. The AML module generates the GUI menus and ties the components together. The INFO database management system stores, maintains, manipulates and reports all spatial and non-spatial attribute data relevant to the DSS.

Modeling System: The modeling system consists of environmental models and an economic model. Two environmental models will be incorporated into WAMADSS: Agricultural Non-Point Source Pollution model (AGNPS) and the Soil and Water Assessment Tool (SWAT). AGNPS simulates erosion, sediment, runoff, and nutrient (nitrogen and phosphorus) transport from agricultural watersheds for individual storm events. SWAT is a continuous daily time-step model which simulates the impacts of alternative land use management practices on surface and ground water, sediment and agricultural chemical yields in ungaged watersheds. SWAT is capable of simulating pesticide transport by runoff, percolate, soil evaporation, and sediment. The economic model evaluates the effects of a particular spatial configuration of LUMAPs on annualized net returns at the field, farm and watershed scales. A spatial configuration refers to the LUMAPs applied to each and every field in the watershed as specified by the user(s). WAMADSS calculates annualized net returns on an acre, field and watershed basis using the Cost and Returns Estimator (CARE). The spatial input data needed to calculate annualized net returns include: set-aside requirement, total acreage per field, planted acreage per field (total acreage times proportion planted), initial crop yields and cost of production per acre. Cost of production is estimated based on crop yield, LUMAP and average costs of farm labor, fertilizer, pesticides, fuel and machinery/equipment.

Graphical User Interface: A GUI provides the user with access to the GIS and modeling system. The GUI contains menus which allow the user to select LUMAPs, parameters and evaluation criteria needed to run the environmental and economic models in WAMADSS. WAMADSS permits the end user to modify land use activities by prompting the user through a series of menus which are used to update the parameters for the selected LUMAPs. A menu provides an interactive interface for entering all the parameters needed to execute a complex operation. The user provides information (filling in blanks, checking boxes or answering questions) by interacting with visual objects called widgets.

All the parameters required for the economic and environmental models are stored as relational tables and accessed through the GUI. Some parameters are based on physical attributes extracted from the various layers (hypsography, landuse, soils, hydrology) while other parameters are based on input elicited from the user via the GUI. WAMADSS' open architecture and modular framework supports the refinement, addition and interfacing of new components. The environmental models utilize spatial and non-spatial (GUI) input data stored in the GIS. Input data for the economic model is supplied by the user via the GUI. WAMADSS allows the user to specify the criteria used to evaluate watershed management plans. Based on the results of WAMADSS, the user can modify the LUMAPs until a desired management plan is achieved. The study area is Goodwater Creek watershed which is located in Boone and Audrain counties, Missouri.

Yan Zhou, Research Associate
Chris Fulcher, Associate Director and Research Associate
Tony Prato, Director and Professor
Center for Agricultural, Resource and Environmental Systems
200 Mumford Hall
University of Missouri - Columbia

Columbia, MO 65211
Telephone: (573) 882-6534
Fax: (573) 882-3958
E-mail: zhou@cares.missouri.edu

David T. Hansen

Metadata and Standards - Communicating between Disciplines in the Encounter between GIS and Environmental Modeling

---

Metadata and Information Supporting Environmental Modeling for Site Specific Applications

Providing Data Quality Information

Metadata or data about data is intended to provide sufficient information that a potential user of the data can make an informed decision about the application and use of the data set. For data sets developed in Geographic Information Systems (GIS) the **Content Standards for Digital Geospatial Metadata** describes the information that should be provided with the digital geospatial data sets (Federal Geographic Data Committee).

Major sections of elements in the standards related to the application and use of the spatial data set include:

Identification Information
: This section describes the intended use of the data set and any known limitations in its use. It provides an opportunity to cite information and other documents related to the data set.

Data Quality Information
: This provides an assesment of data quality for both the spatial coordinates of the features represented and the attributes that are assigned to those features. In addition, it provides information on the lineage, logical consistency, and completeness of the data.

Entity and Attribute Information
: This section provides a description of the entity and attribute structure of the data set. This includes the definition of the attributes, the range of known or permitted values, and the source of the definitions for the attributes and their values.

Scientists and engineers are concerned about the proper application and use of the data and models that they prepare for environmental applications. Information collected in the development of environmental models is documented and described in a variety of forms following standard procedures for the particular disciplines that are involved. Having access to this information enhances the application and use of the data sets.

However, this information is often difficult to capture and to include with the metadata prepared for the digital geospatial data sets. It is often contained in separate documents. These documents can consist of other digital files, hard copy reports, maps, aerial photography, photos or video of sites and sampling procedures, etc. Even where documentation is primarily concerned with the spatial coordinate structure, there can be situations where this information is more easily handled in separate documents. The **Content Standards for Digital Geospatial Metadata** provide the opportunity to cite and where appropriate link to digital versions of these related documents. The metadata for the geospatial data set can identify contacts for this information when other documentation is unavailable.

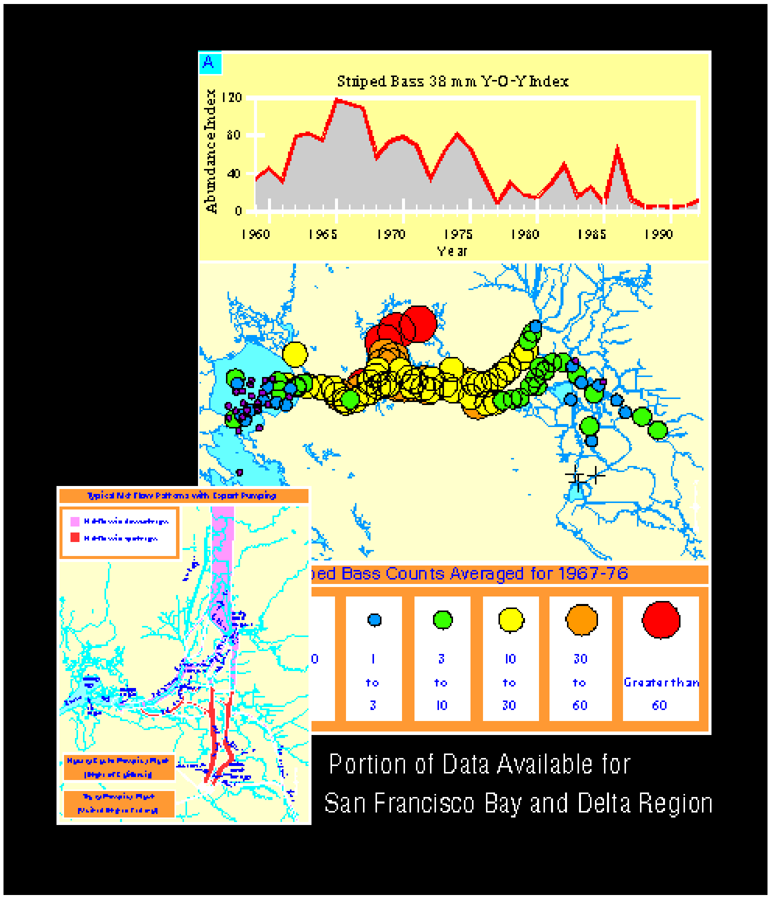
Examples Illustrating the Complexity of Documenting Environmental Modeling Applications

The following examples are used to illustrate the complexity of documenting environmental modeling applications:

- The complexity of fully describing the attribute data for the Interagency Ecological Studies Program for the Sacramento - San Joaquin Estuary, California.
- The amount of information required to fully document the application of ground-water models to a site specific problem in a GIS interface.
- The complexity of documenting the coordinate control structure for some digital spatial data sets where that structure is to be adjusted.

Environmental Data for the Interagency Ecological Studies Program

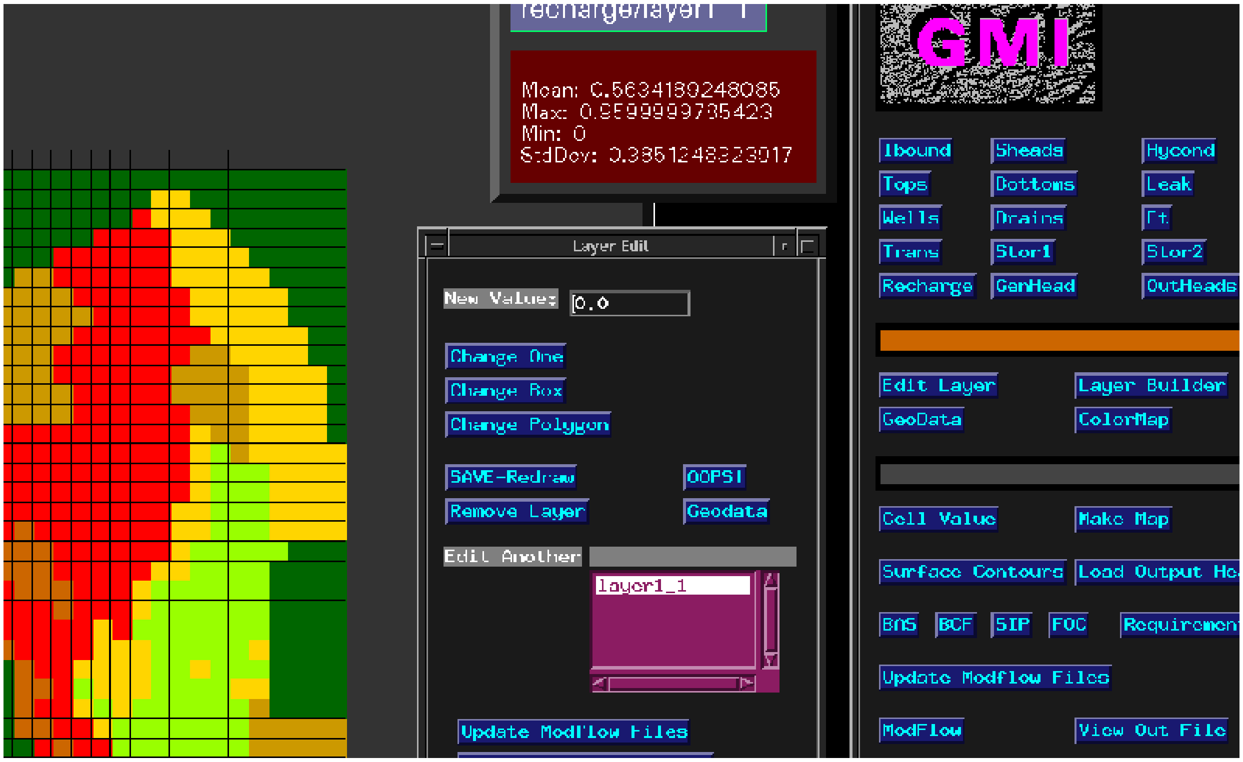
The Sacramento - San Joaquin estuary is a major aquatic and wildlife habitat at the head of San Francisco Bay. This estuary is also a primary supply for public and irrigation water for northern and southern California. GIS has actively been used in the analysis of trends in water quality and biota populations as well as modeling of those trends under different flow regimes.



The sampling of biota and modeling of the system for various species is particularly complex. Types of data, sampling methods, and analysis procedures vary depending on time period and the media being studied. Data include several life stages for a variety of organisms collected at different time periods under a variety of flow regimes. The appropriate use and application of this data requires access to individuals associated with the studies or to documents describing the studies. These documents consist of an extensive set of reports, studies, and articles prepared by the separate agencies involved in these studies. They are prepared following standards or procedures in place at the separate agencies. It would be redundant and impractical to include all of this information in the metadata for the digital geospatial data sets. Metadata for the geospatial data sets provide the opportunity to cite these reports and where available provide links to digital versions of these documents. Where these documents are not readily available, the metadata can provide names and addresses of individuals to contact for additional information (Hansen, 1996, ASTM STP 1279).

GIS Interface for Ground-Water Models in the San Joaquin Valley

The San Joaquin Valley of California is one of the major agricultural regions in the United States. The valley contains fertile soils and has a long growing season that permits the production of a wide variety of crops. The climate is arid to semiarid and irrigation requires drainage for some argicultural areas. This agricultural drainage water has posed water quality problems for areas both within the valley and the San Francisco Bay. Ground- water models have been used as a tool to assist in the evaluation of alternative methods for managing agricultural drainage related problems.



The existing base will be migrated over to a new base that more accurately represents the location of the quarter - quarter section corners. This will improve the representation of aerial features within all of the data sets. The existing map projection system which is based on the 1927 North American Datum will also be examined for conversion to a new datum. This transformation of the underlying coordinate structure of the data sets is described in a series of reports.

The adjustment of coordinate values for individual GIS data themes will require documentation of the processing steps down to the feature or entity level. Reporting will be required of changes to values of area and length for the geospatial features. This adjustment in the spatial component of all features in the data set will generate voluminous metadata. This is most easily handled in separate reports of the adjustment process that can be referenced in the metadata for the individual geospatial data sets.

The Role of Standards

For many disciplines, standards have been prepared or are under active review and development by numerous organizations. Standards provide a common lexicon of terms, a uniform set of procedures, and a common content for reporting on the modeling effort. GIS has entered the standardization process in the development of SDTS and with the **Content Standards for Digital Geospatial Metadata**. These standards identify metadata elements that are mandatory or optional but no consensus guides have yet been developed for the application of these standards. Moellering identifies several areas where the GIS community should continue to persue further development of consensus. These include:

- Methods for reporting on data quality particularly as it relates to the use and application of the data,
- Means for describing extensive lineage or processing histories of spatial data,
- Identification of common entity definitions for base features.

(Moellering, 1994). Buttenfield identifies the need for the development of common terms for spatial data quality and accepted methods for graphically representing data quality (Buttenfield, 1993). Standards developed within the GIS community need to complement the standards commonly accepted by other disciplines in the environmental modeling community. This will assist in providing access to the information required for the proper application and use of environmental modeling data.

Independent standards organizations provide a neutral forum for the development of consensus standards involving a variety of public and private interests. ASTM is one such independent standards organization. The **Content Standards for Digital Geospatial Metadata** are presently under final review for adoption as ASTM **Content Specifications for Digital Geospatial Metadata** (D5714). The independent review process has enhanced these standards by the assignment of unique tag names and unique tag values to individual metadata elements. It has active subcommittees involved in the standards process on environmental metadata, GIS, remote sensing, site characterization for environmental purposes, geostatistics, water quality, and ground-water modeling. In addition to ASTM, there are several other standards organizations such as:

- American Chemcial Society
- American National Standards Institute
- American Society of Agricultural Engineers
- American Water Works Association
- National Information Standards Organization
- National Institute of Standards and Technology

I encourage you to become active in any of these standards organizations.

References

ASTM. 1993. **Standard Guide for Application of a Ground-Water Flow Model to a Site-Specific Problem**. D 5447 American Society for Testing and Materials. Philadelphia

Buttenfield, Barbara P. 1993. **Representing Data Quality**. Cartographica Vol. 30 No. 2-3 pp 1-7.

Federal Geogrphic Data Committee. 1994. **Content Standards for Digital Geospatial Metadata**. Washington, D.C.:June 8, 1994

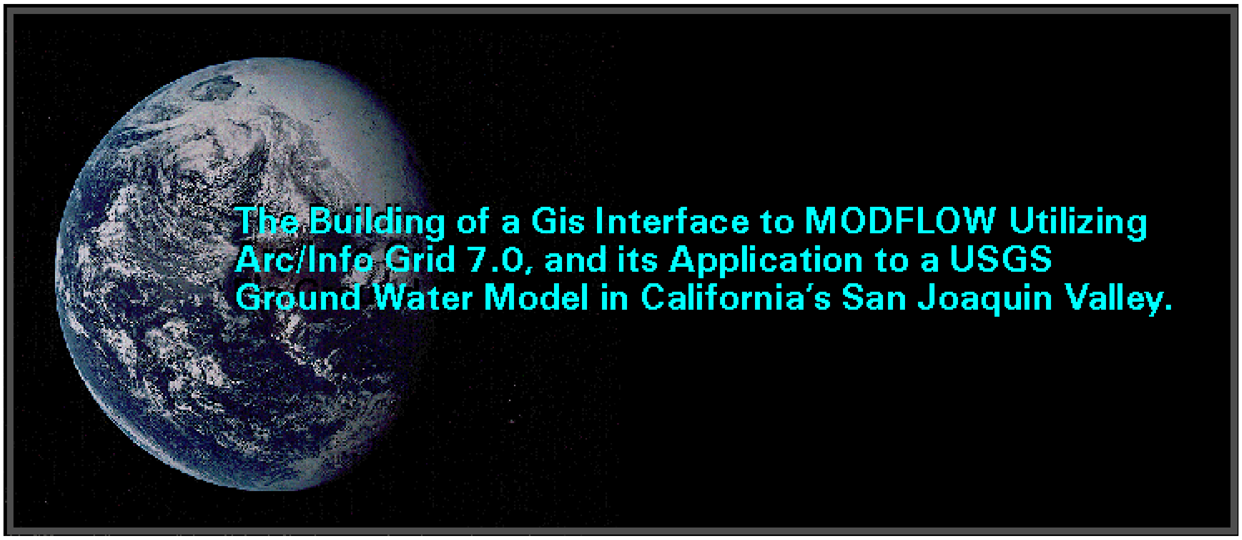
Hansen, David T., 1996. **Application of Content Standards for Digital Geospatial Metadata to Laboratory Data and the Site Characterization Process**.International Symposium on Remote Sensing and GIS for Site Characterization, ASTM STP 1279,. V.N. Singhroy, D.D. Nebert, and A.I. Johnson, Eds.; American Society for Testing and Materials, Philadephia.

Hansen, David T., 1996. **Documentation of Ground-Water Models in Relationship to National Digital Geospatial Metadata Standards.** .Subsurface Fluid-Flow (Ground-Water and Vadose Zone) Modeling , ASTM STP 1288,. Joseph D. Ritchey and James O. Rumbaugh Eds.; American Society for Testing and Materials, Philadephia.

Moellering, Harold. 1994. **Continuing Research Needs Resulting from the SDTS Development Effort**. Cartography and Geographic Information Systems., Vol. 21 No. 3 pp 180-189.

David T. Hansen, GIS Specialist / Soil Scientist
Mid Pacific Region
U.S. Bureau of Reclamation
Department of Interior
2800 Cottage Way
Sacramento, CA USA 95825-1898
Telephone: (916) 979-2418
Electronic Mail: dhansen@mpgis7.mp.usbr.gov

Thomas Heinzer - Michael Sebhat - William Greer



The U. S. Bureau of Reclamation is responsible for natural resource management, including many aspects of water storage and delivery. Five years ago, it was envisioned at Reclamation that the most efficient way to handle complex models was not only to pre- and post-process model data in GIS systems, but to have the model data RESIDE in GIS data structures, and to 'talk' to the model directly through the GIS sytems via graphical user interfaces (GUIs). Using this approach, the modeler can directly 'see' and readily modify model data graphically, with the aid of ancillary data sources, thus greatly facilitating data verification and editing.

Over the years, Reclamation has developed Geographic Information System (GIS) interfaces for both surface and ground water models, with much focus in the Central Valley, of California. Currently, Reclamation is utilizing an internally developed, GIS based software named 'Grid/MODFLOW Interface', or GMI(1), to manage a ground water model based on the USGS numerical code MODFLOW(2). This software runs on a Unix/ARC/INFO(3) platform, and takes advantage of the raster, or 'grid' capabilities of GIS sytems to store and manage model data.

One application of interest is the interface of GMI to modify a previously generated U.S.G.S. MODFLOW model known as the Fio(4) model. Certain areas of the Central San Joaquin Valley have sustained relatively high water tables, resulting in drainage and water quality concerns. This application will be used to evaluate drain effluents as a function of differing cropping patterns and irrigation scenarios. If successful, the application will be linked to a real time operations model simulating adjacent canal flow and water quality, to determine optimal times to release drain effuent into the canals or rivers to comply with water quality regulations.

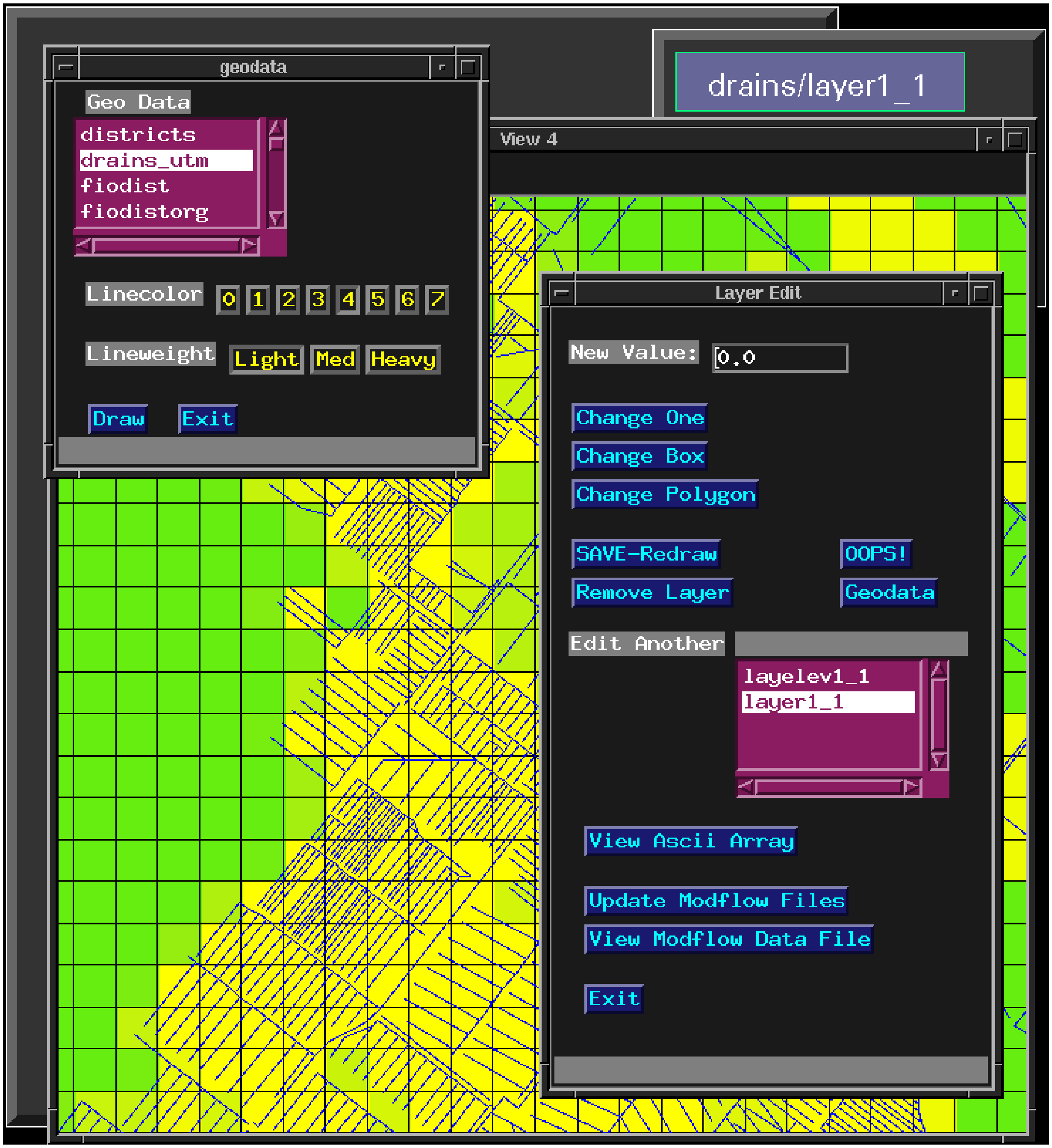
The original model has been extended, undergone mesh refinement and data modification, and has been changed from a steady state to transient model. Irrigation and evapotranspiration rates vary monthly, and the application is used to determine drainage budgets at certain points of the drain network. Since the irrigation rates were associated with irregular 1:24000 resolution field polygons, GIS techniques were used to distribute these rates to the uniform finite difference

mesh using area weighting.

This poster presentation illustrates the use of GMI and some of the advantages of this type of interface. Of particular interest is:

* **All of the model data reside in raster or 'grid' layers that can be viewed and edited. The MODFLOW ready input files are automatically generated, transparent to the user.**
* **Model input and output data may be readily viewed simultaneously with ancillary vector data, thus facilitating decision making and troubleshooting.**
* **Map algebra(5) may be used to manipulate model data layers. This allows a model layer's cells to be set to a number or some function of another data layerwhich permits the rapid evaluation of management scenarios that would otherwise be monumental undertakings.**

Below is an example of a view of a modeled area using GMI. The yellow areas represent greater drain conductances, and the blue lines are the known drain locations in vector format.



For additional information contact Michael Sebhat or Thomas Heinzer at:

U.S. Bureau of Reclamation
The MPGIS Service Center
2800 Cottage Way
Sacramento, CA. 95825
(916) 979-2441

References:

(1) THE BUILDING OF A GIS INTERFACE TO MODFLOW UTILIZING ARC/INFO GRID 7.0; Thomas Heinzer, Michael Sebhat, William Greer, David Hansen; Paper presented at the 1995 Environmental Systems Research Institute National User Conference.

(2) A MODULAR THREE-DIMENSIONAL FINITE-DIFFERENCE GROUND-WATER FLOW MODEL; Michael G McDonald and Arlen W. Harbaugh, U.S.G.S. Open-FIle Report 83-875

(3) ARC/INFO is a Geographic Information System software produced by Environmental Systems Research Institute, Redlands, California.

(4) CALCULATION OF A WATER BUDGET AND DELINEATION OF CONTRIBUTING SOURCES TO DRAIN FLOWS IN THE WESTERN SAN JOAQUIN VALLEY, CALIFORNIA; John Fio, USGS Open-File Report 94-95

(5) Map algebra is language designed to perform logical expressions on ARC/INFO GRID data layers.

Louis R. Iverson and **Anantha M. Prasad**

Modeling Present and Potential Future Tree Importance Values in the Eastern United States

Abstract

Regression tree analysis (RTA), when integrated with geographic information systems (GIS), allows for improved interpretation of species-environment interactions and prediction of current and potential future species distributions. We have been assessing the environmental factors associated with current tree species ranges to better understand their potential to survive and/or migrate under a changed climate or other disturbance. We have collected, summarized, and analyzed data for climate, soils, land use (including the spatial configuration among land use types), socio-economic factors, and species assemblages for over 2100 counties east of the 100th meridian. Forest Inventory Analysis (FIA) data for over 100,000 forested plots in the East provided the tree species range and importance values information for 103 species of trees. Regression tree analysis in S-PLUS was used to devise prediction rules from current species-environment relationships, which were then used to replicate the current distribution as well as predict the future potential distributions under two scenarios of climate change with 2xCO_2. Validation measures prove the utility of the RTA modelling approach for mapping current tree importance values across large areas, leading to increased confidence in the future predictions. Although these future predictions do not address the fate of migrating species through fragmented landscapes, they do give an idea of the basic envelope to which the species may be adapted should no restrictions to migration apply. Graphical outputs from RTA, combined with the predicted tree species distribution maps in GIS, provide a powerful means of understanding the relationships among various factors associated with tree species distributions.

Introduction

Environmental drivers, as modified by disturbance processes, generally control the distribution of tree species, and these relationships are increasingly being borne out. Within a region, species vary primarily due to regional climatic factors, whereas variations in terrain, soil, and land-use history factor principally in more local studies. Geographic information systems (GIS) allow predictive mapping of vegetation based on the species-environment relationships. Evidence is also mounting that anthropogenic greenhouse gases and sulphate aerosols are related to a general warming trend on the planet (MacCracken 1995, Wigley 1995). Various global circulation models (GCM) predict 2-6 deg C temperature increases as a result of doubling atmospheric CO_2. This suggests a potential for major changes in the earth's living systems, including temperate forests.

Regression tree analysis (RTA) is a relatively new techinique being used in the ecological sciences which uses repeated resampling of the data to develop empirical relationships between response and predictor variables, rather than the more restrictive distributional assumptions in classical regression functions. This alternative modeling approach creates models which are fitted by binary recursive partitioning whereby a dataset is successively split into increasingly homogeneous subsets which elucidate relationships between predictor and response variables (Clark and Pregibon 1992). The RTA approach seems appropriate for predicting landscape-level distributions of species from environmental data. Its use has grown with that of geographic information systems, which allow model outputs to be readily mapped across landscapes. There are few ecological examples of the use of RTA (e.g., Davis and Goetz 1990, Michaelsen et al. 1994). In this study, we use RTA to evaluate the relationship of 63 environmental variables to 103 eastern tree importance values, and then use the derived relationships to predict their present and potential future ranges.

Methods

Data were extracted from several sources for the land east of the 100th meridian. Counties were chosen as the mapping unit because they are the reporting unit for many sources and are similarly sized in the East. A total of 67 environmental/landuse/socioeconomic variables were used for each of 2100+ counties. Data collected and transformed to county-level data included:

1. Tree ranges and importance values, as calculated from over 100,000 forest inventory and analysis plots assessed by the USDA Forest Service (Hansen et al. 1992);
2. Climatic factors, as obtained from the USEPA (1993);
3. Soil factors, from the State Soil Geographic Data Base (STATSGO) data base (Soil Conservation Service 1991);

4. Land use/land cover, from the GEOECOLOGY data base of Oak Ridge National Laboratory (Olson et al. 1980) and AVHRR-derived forest vegetation classes from the USDA Forest Service (1993);
5. Socioeconomic factors, from ArcData (ESRI,1992);
6. Elevation, derived from 1:250,000 USGS 3 arc-second data (USGS 1987); and
7. Landscape pattern, as calculated on the AVHRR forest cover map using Fragstats (McGarigal and Marks 1994).

Climate data included projected outputs from two scenarios of equilibrium climate under 2xCO2: the GFDL (Geophysical Fluid Dynamics Laboratory, Wetherald and Manabe 1988), and GISS (Goddard Institute of Space Studies, Hansen et al. 1988).

Regression trees were generated in S-PLUS (Statistical Sciences 1994) for 103 tree species in the FIA database (Figure 1). Species importance value (based on basal area and number of stems) was the response variable, along with the 67 predictor variables mentioned above. The regression trees were generated, with a random selection of 80% of the data, and used to predict the importance value for each species by county. These predictions were then output to Arc/Info for mapping, using Unix scripts and Arc macros. To evaluate the model outputs, a comparison of predicted current and actual (FIA) distributions were made using correlation, verification, and validation processes.

Once the regression trees were generated, they could not only be used to generate predictive maps of current distributions, but also potential future distributions under a scenario of a changed climate. For this, we swapped predicted future climate variables, according to the GFDL and GISS models, for the current county estimates of the climatic variables (Figure 1).

Obviously, a study of this type carries a suite of assumptions and limitations which must be stated. They fall into three categories: data inputs (e.g., overlay of multiple GIS layers, uncertainty of input variables, especially GCM outputs), analysis (e.g., correlation among variables) , and biology (e.g., no consideration of changes in competition or water use efficiency).

Results and Discussion

Predicted current distributions match FIA data quite well for most species. For example, in Quercus falcata var. falcata (southern red oak, Figure 2), the FIA data recorded the species in a total of 840 counties. Total classification accuracy was 76% on the entire data set and 74% on the 20% validation data set. If only considering error of omission (the RTA model predicted no target species when the FIA sampling recorded it), the accuracies are 91 and 92%, respectively. The correlation between actual and predicted importance values was 0.87. The species generally grows on upland sites over a wide variety of soil conditions (Belanger 1990). The county-level of resolution is therefore adequate to capture the major environmental variables driving its distribution, and conditions represented by county averages are adequate to model the species. In general, the more specialized the species is with respect to edaphic conditions, the less accuracy in the RTA model predictions, because county-resolution data would not be expected to consistently capture the appropriate information for the RTA process.

Projected species distributions, following equilibrium of predicted climate changes, show major shifts for many species. For example, southern red oak is projected to shift its northern limit by nearly 400 km to the north (Figure 2). Because of differential shifts among species in area and importance, forest communities would also be expected to differentially change in composition. Projections also show the predicted changes in composition for any county, as well as the spatial trends in overall tree species richness.

Conclusions

1. RTA is a valuable tool to understand species-environment relations.
2. RTA can be used to predict potential migrations of trees under a 2xCO2 climate scenario. It assumes the species will migrate as needed with no restrictions. By operating at a species level, however, it also gives an indication of potential changes in community dynamics and biodiversity.
3. Historic rates of migration (~10-50 km/century) will not likely occur with the current fragmented habitat (Schwartz 1993). Even at historic rates, many species would not reach the potential distributions predicted here within the next century. What would happen to the species that cannot keep pace with the climate forcing?
4. Current investigations with colleagues are assessing more realistic migration scenarios based on species regeneration characteristics and actual landscapes.

Acknowledgments

Sincere thanks are due all the people that provided data for this effort, and to the USDA Forest Service, Northern Global Change Program (R. Birdsey, Program Manager) for their support.

References

Belanger, R. P. 1990. Quercus falcata Michx. var. falcata. Pages 640-644 in Burns. R.M. and B. H. Honkala, Coordinators. Silvics of North America. Volume 2, Hardwoods. USDA Forest Service, Agriculture Handbook 654, Washington, DC.

Clark, L. A. and Pergibon, D. 1992. Tree-based models. Pages 377-419 in T. J. Hastie. Statistical Models in S. Wadsworth, Pacific Grove, California.

Davis, F. W. and S. Goetz. 1990. Modeling vegetation pattern using digital terrain data. Landscape Ecology 4:69-80.

Environmental Systems Research Institute. 1992. ArcUSA 1:2M, User's guide and data reference. Environmental Systems Research Institute, Redlands, CA.

Hansen M H, T. Frieswyk, J. F. Glover, and J. F. Kelly. 1992. The eastwide forest inventory data base: users manual. General Technical Report NC-151, USDA Forest Service, North Central Forest Experiment Station. St. Paul, MN. 48 pp.

Hansen, J., I. Fung, A. Lacis, D. Rind, S. Lebedeff, and R. Ruedy. 1988. Global climate changes as forecast by Goddard Institute for Space Studies three- dimensional model. Journal of Geophysical Research 93:9341-9364.

McGarigal, K. and B. J. Marks. 1994. Fragstats. Version 2.0. Forest Science Department, Oregon State University, Corvallis, OR.

MacCracken, M. C. 1995. The evidence mounts up. Nature 376:645-646. Michaelsen, J., D. S. Schimel, M.A. Friedl, F. W. Davis, and R.C. Dubayah. 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. J. of Vegetation Science 5:673-686.

Olson, R. J., C. J. Emerson, and M. K. Nungesser. 1980. Geoecology: a county-level environmental data base for the conterminous United States. Oak Ridge National Laboratory Environmental Sciences Division Publication No. 1537, Oak Ridge, TN.

Schwartz, M. W. 1993. Modeling the effect of habitat loss on potential rates of range change for trees in response to global warming. Biodiversity and Conservation 2:51-61.

Soil Conservation Service. 1991. State soil geographic data base (STATSGO) data users guide. Miscellaneous Publication 1492, USDA Soil Conservation Service. Washington, D.C. 88 pp.

Statistical Sciences. 1994. S-PLUS Guide to Statistical and Mathematical Analysis, vers. 3.3. StatSci, Seattle, WA.

USDA Forest Service. 1993. Forest type groups of the United States. Map produced by Z. Zhu, D. L. Evans, and K, Winterberger, Southern Forest Experiment Station, Starkville, MS.

USEPA. 1993. EPA-Corvallis model-derived climate database and 2xCO2 predictions for long-term mean monthly temperature, vapor pressure, wind velocity and potential evapotranspiration from the Regional Water Balance Model and precipitation from the PRISM model, for the conterminous United States. Digital raster data on a 10 x 10 km, 470x295 Albers Equal Area grid, in "Image Processing Workbench" format. USEPA Environmental Research Laboratory, Corvallis, OR.

US Geological Survey. 1987. Digital elevation models: U.S. Geological Survey Data Users Guide 5. US Geological Survey, Reston, VA.

Wetherald, R. T. and S. Manabe. 1988. Cloud feedback processes in a general circulation model. Journal of Atmospheric Science 45:1397-1415.

Wigley, T. M. L. 1995. A successful prediction? Nature 376:463-464.

Figure 1

Flow diagram for RTA analysis.

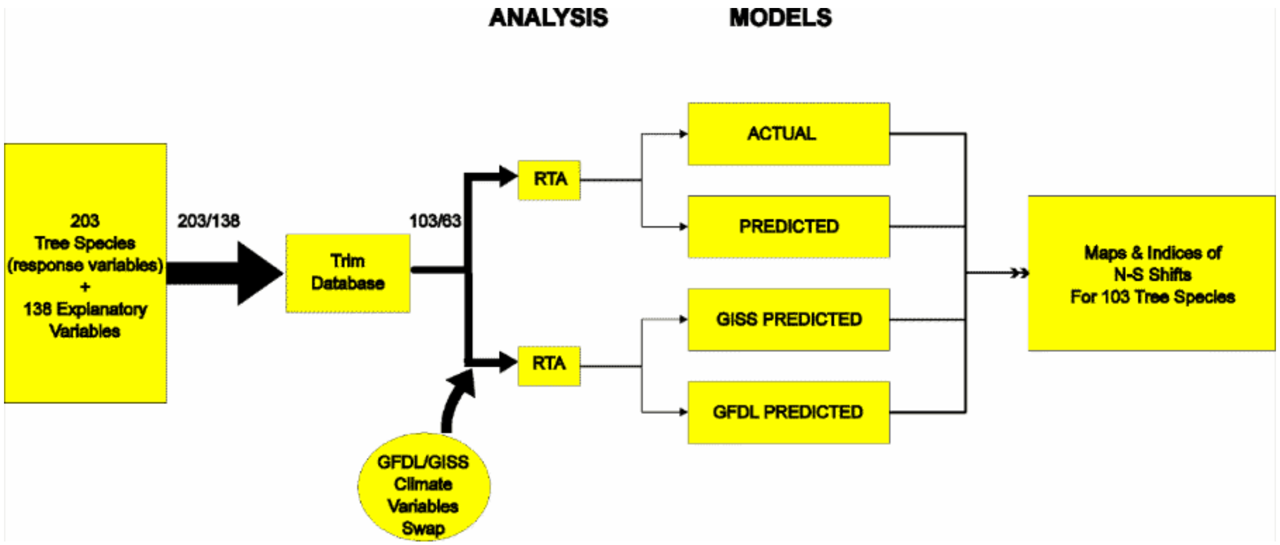
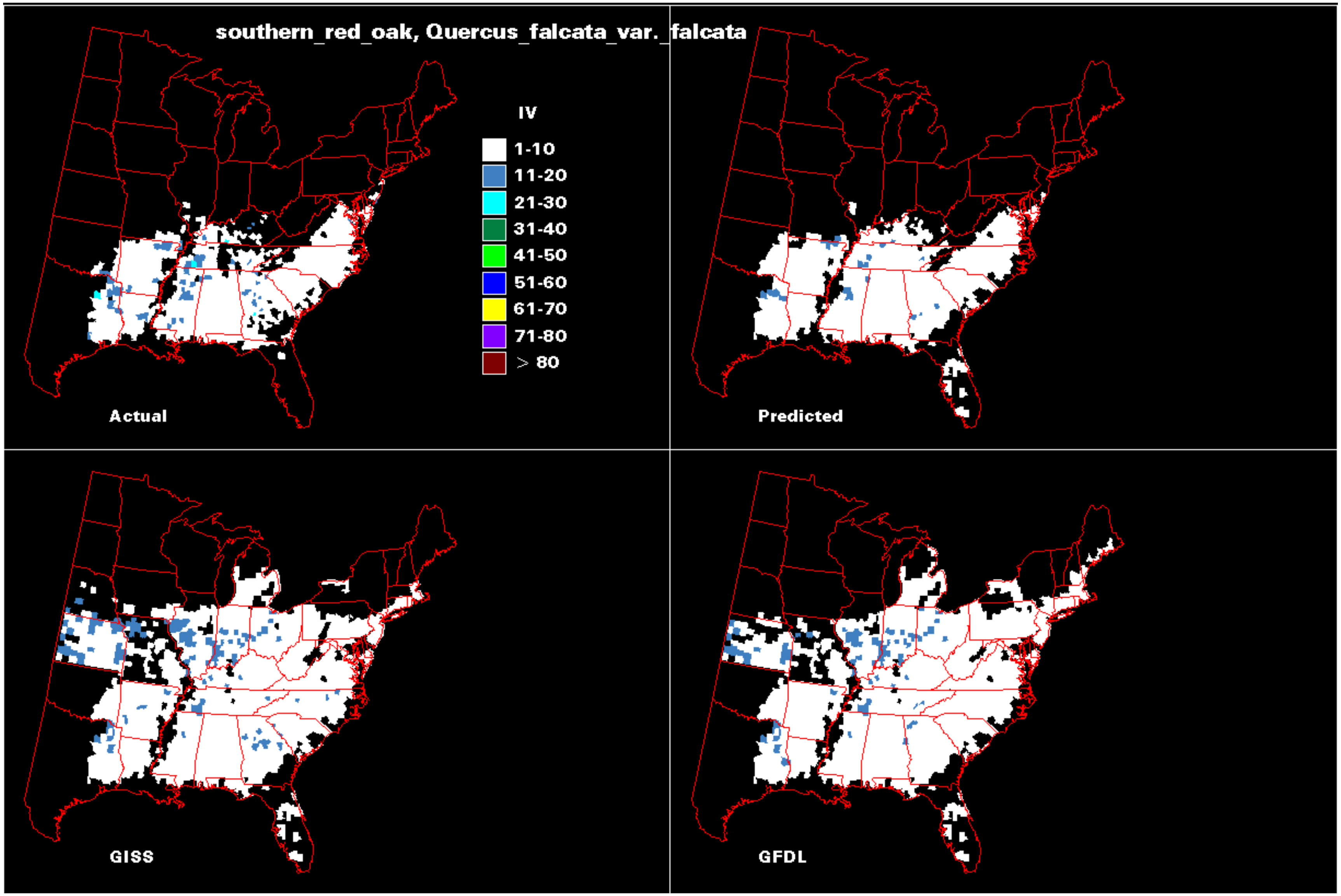


Figure 2

Example model outputs for Quercus falcata var. falcata (southern red oak), including:

1. actual county importance values as calculated from FIA data;
2. predicted current importance values from the RTA model;
3. predicted potential future importance values after climate change according to the GISS GCM;
4. predicted potential future importance values after climate change according to the GFDL GCM.



Louis R. Iverson
Northeastern Forest Experiment Station

359 Main Road
Delaware, Ohio 43015
614-368-0097 (voice)
614-368-0152 (fax)
iverson@trees.neusfs4153.gov

Anantha M. Prasad
Northeastern Forest Experiment Station
359 Main Road
Delaware, Ohio 43015
614-368-0103 (voice)
614-368-0152 (fax)
prasad@trees.neusfs4153.gov

Stephen R. Kessell

What are the educational requirements for integrating GIS and environmental modeling?

Western Australia's Curtin University of Technology was one of the first universities to offer a Bachelor of Science degree in GIS. The degree involves a significant amount of study in mathematics, computing and spatial analysis subjects, as well as an "applications minor" (which is often taken in an environmental science area). This poster presentation explores the rationale and content of this degree program.

The School of Computing at Curtin was reorganised in 1992 into two separate departments: GIS and computer science. There were several reasons for doing this, with the two main ones being to:

- provide a clear and visible presence of GIS at the university
- offer a new and possibly unique Bachelor of Science degree in GIS

We previously had offered a Graduate Diploma in GIS. The School's major research thrust is in the area of spatial and environmental computing, bringing together research tools from GIS, AI, graphics, visualisation, database design, parallel processing, etc. So why a separate department and degree?

We chose to offer a "hard science" degree in GIS, with all students receiving solid grounding in calculus, statistics, computer science (theory as well as programming) and written and oral communications skills, while they concurrently obtained their specialised training/education in GIS. Units such as "spatial statistics", "spatial analysis and modelling" and "geographic data structures" are mandatory subjects.

Thus the degree is very strong in the area of spatial analysis, with obvious applications to the environmental sciences.

Another feature of this degree is that all students must undertake a "minor" which includes at least six additional subjects in an "applications area" (such as environmental science, geophysics, remote sensing, town planning, etc. -- choice is left to the student). A significant number of our students specialise in the environmental fields, encouraged by excellent employment opportunities in this area in Western Australia. This structure thus allows a student to combine both intensiv e "hard science" study in GIS with specialisation in an area such as environmental modeling; it also allows students to pursue a "double degree" if desired.

With the first cohort graduating in 1994, the feedback from employers, and the employment opportunities for students who chose not to go on to higher degrees, suggest that this degree structure is working well.

Since 1992, we have added an honours degree in GIS as well as a Post-Graduate Diploma, Master of Science, and PhD, which provide a range of coursework and research options for further study. There is close collaboration between the "GIS" and "computing" people in all of the research degree. Our experiences to date with these programs will be discussed.

Steve Kessell is Associate Professor and Head of the Department of Geographic Information Systems, School of Computing, Curtin University of Technology, Perth, Western Australia. Trained as a plant ecologist (Amherst College and Cornell), Steve joined Curtin in 1987 to develop the undergraduate and GIS programs in the (then new) computing school. His major interest has evolved from research to curriculum development over the past several years, and he is currently completing a Master of Education degree.

R.Z. Khamitov, V.E. Gvozdev, S.V. Pavlov, A.N. Vasiliev

Environmental Modeling and Prognosis System as a Functional Subsystem of Ecological Information Management System of the Republic of Bashkortostan

Ecological Information Management System (EIMS) is being created in the Republic of Bashkortostan. The system is intended for prompt providing government authorities with comprehensive and reliable information on environmental state and quality. The EIMS as an information system unites into a single whole information systems of all government organizations in the Republic of Bashkortostan territory. The EIMS uses official information of state control organizations, current information received from automatic stations of atmospheric air and surface waters monitoring, scientific research data. All this information is being stored by various organizations in their local data bases (LDB). The EIMS ensures logical integration of LDB into a single distributed data base (DDB).

A system of environmental modeling and prognosis within the EIMS is considered as an instrument providing for higher validity of decisions taken in the process of environmental management in the region.

Mathematical modeling and prognosis system alongside with other information sources (geoinformation systems GIS, distributed data bases of environmental character DDB, satellite information processing system SIPS for treatment of environmental information) make an information basis for functioning of the system for providing variants of management decisions on environmental safety, the system for adopted decisions realization, the system for evaluating efficiency of fulfilled measures on environmental safety. (Figure 1)

Figure 1
(not available)

Mathematical modeling and prognosis system may be considered, on the one hand, as a means of processing primary information coming from other functional subsystems of integrated information system)GIS, DDB, SIPS), and on the other hand, as a source of primary information for the system of information support for environmental safety activity.

Figure 2
(not available)

This system comprises the following subsystems:

- an executive subsystem intended for computer realization of various mathematical models;

- a supplying subsystem intended for providing the executive subsystem

with required initial information;

- a geoinformation system which serves, on the one hand, as a source of initial data for the executive subsystem, and on the other hand, as an instrument for spatial rearrangement of mathematical modeling results and their representation on maps;

- a user's request adaptation subsystem intended for selection of a mathematical model out of a set of available models in compliance with the task for modeling, information requirements of models and available initial data.

Geoinformational model of the Republic of Bashkortostan (GIM RB) includes topographical maps of the RB territory and applied software programs working with GIS.

GIM RB contains the following electronic maps of the Republic territory:

- the Republic of Bashkortostan territory at the scale of 1 : 1 000 000, singled out from the digital world map ESRI DCW, specified and corrected;

- the Republic of Bashkortostan territory at the scale of 1 : 700 000;

- topographical map of the Republic of Bashkortostan territory at the scale of

1 : 200 000, containing relief, hydrography and hydraulic constructions, populated areas, communication lines and electric power lines, road net, vegetation, pipeline net, administrative division;

- schematic maps of lands usage for some regions of the Republic;

- plans of some large industrial enterprises.

Creation of digital maps is being carried on in ARC/INFO environment for PC and SUN SPARK operation stations and with the use of Easy Trace vectorizer for scanned images.

Applied programs working in GIS environment and developed on the basis of program languages Avenue ArcView 2, Borland Pascal, Delphi, AML ARC/INFO v.7.0.3 include:

- statistic and dynamic models of pollution spreading through water and air;

- definition of areas damaged as a result of accidents at extra hazardous objects;

- definition of optimal routs for emergency means and transport passing via transport net to the objects in emergency zone;

- definition of the range of reaching for emergency means while passing along transport net;

- definition of natural flood zones or flooding as a result of building hydrotechnical constructions;

- definition of ways of draining off liquids in case of their spreading or pipelines breaks.

The modeling system is a distributed system. Within the EIMS

it is considered as a logical integration of local modeling systems existing in various organizations.

GIS occupies the central position in the distributed system of modeling. First of all this is due to the fact that local modeling systems use GIS as a source of initial data and as an instrument of modeling results representation , secondly, due to the fact that effective use of GIS requires expensive software and equipment, as well as specialists thoroughly trained in the field of information technologies.

The main principles of developing information technologies for matching modeling results with GIS are as follows:

- in a distributed modeling system a multi-task and multi-user mode is to be maintained;

- a possibility of connecting previously created modeling programs,
designed for independent use, to the modeling system is to be secured;

- different level of software and equipment quality in different organizations is to be taken into consideration; changes in software and equipment should not influence serviceability of the distributed modeling system;

- local modeling systems are to maintain their serviceability regardless of other subsystems serviceability.

At present information technologies for solving the following tasks have been developed: simulating air pollution as a result of industrial enterprises activity, surface waters pollution by waste waters; revealing the main sources of diffusion pollution of surface waters, prognosis of possible consequences of accidents at petrochemical plants, pipelines and so on. A geoinformation model of the Republic of Bashkortostan territory and its main cities has been created. ARC/INFO has been used as GIS software.

Author information:

Rustem Z Khamitov, Minister, Ph.D.
Ministry of Emergency Situations and Environmental Safety

Vladimir E. Gvozdev, Chief of Laboratory for System Analysis and Environmental Modeling, Ph.D., Assistant Professor
Institute for Problems of Applied Ecology and Natural Resources Use

Sergei V. Pavlov, Deputy Director, Ph.D., Assistant Professor
Institute for Problems of Applied Ecology and Natural Resources Use

Andrei N. Vasiliev, Acting Chief of Laboratory for Automatic Processing of Cartographic Information
Institute for Problems of Applied Ecology and Natural Resources Use

Post address: 12/1, 8 Marta St., 450005, Ufa, Bashkortostan, Russia

Tel(fax): (3472)28-75-90

E-mail: root@ippeprb.bashkiria.su

USING GIS AND COUPLED MODELS FOR UNDERSTANDING FOREST ECOSYSTEM DYNAMICS

E. Levine[1], R. Knox[1], J. Ranson[1], J. Weishampel[1], J. Smith[1], G. Sun[1], A. Friend[2], S. Fifer[1], and D. Williams[1]
[1]NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771 USA
[2]Institute of Terrestrial Ecology, Edinburgh Research Station, Scotland, UK

The Forest Ecosystem Dynamics (FED) Project involves the development and integration of models to understand soil, vegetation, and radiation dynamics in northern forest ecosystems. Through the use of simulation models, remote sensing, field investigation, and GIS, the vegetation, soil, and energy components within the northern forest are being investigated, and their responses to global change or other disturbance are being explored and quantified.

In the first phases of the FED project, data collection and model development has focused on the International Paper's Northern Experimental Forest (NEF) at Howland, Maine. Intensive field measurements coinciding with aircraft and satellite overflights, as well as ancillary data sets were obtained and incorporated into the FED GIS. Models of soil physics and snow dynamics, forest growth and tree physiological processes, and thermal dynamics were developed and linked using a modeling environment. The GIS was used to provide driving variables for the models, to validate model predictions, and to identify areas in the field requiring more intensive study. Preliminary results using linked models have been compared to field data and have identified areas for additional work. Similar modeling of sites in the boreal forest region of Canada, which are part of the Boreal Ecosystem-Atmosphere Study (BOREAS), is also underway.

GISMO: A user interface for linking GIS with the EPIC simulation model

Martin, T.C. and Neiman, H.

Crop/soil simulation models require detailed, temporal data for a single point in the landscape, with multiple iterations of many simultaneous equations. Researchers often wish to run multiple simulations for each of the varying conditions within a landscape, but map data structures and map calculator algorithms of GIS packages are not well suited to this kind of modeling. A third software package would be useful, to manage both the simulation model and the GIS, translating and linking the input and output requirements of both. GISMO provides this kind of linkage. GISMO will be introduced and data base and software design issues and decisions will be discussed.

R. Gavin McGhie, Karen Kline, John E. Estes

Creation of a Managed Areas GIS Database of the Conterminous United States for use in Ecosystem Analysis of Managed Versus Unmanaged Areas

---

As the world's population grows, human impact on the earth's ecosystems continues to expand. For this reason, selected areas need to be set aside and certain uses precluded so that at least portions of these ecosystems can remain in their natural state. Selected regions of the earth have been put into different types of reserves and protected areas. Many areas have also been set aside and managed for military, national security or cultural reasons. As a result of restraints placed on their uses, these latter areas often become valuable in protecting biotic species by default. Currently, little data is available about the location and extent of these managed areas especially for studies of large areas. Those data which do exist are mostly in analogue format, and digital conversion is expensive and time consuming.

There is an urgent need for these data. Protected areas data sets covering the entire globe are useful in the assessment of current ecosystem protection and to make recommendations regarding future plans. Much attention has been focused on the destruction of tropical rainforests in developing nations, but ecosystem status in developed nations, including the United States, must also be analyzed. To help meet these needs, a digital Geographic Information System (GIS) database of managed areas was developed for the conterminous United States.

This digital database will allow ecosystems to be more easily and accurately studied using a variety of analytic methods. Digital overlay of managed areas with other databases, such as ecoregions or remotely sensed vegetation data, can be carried out quickly and easily. Statistics may then be computed regarding ecosystem or vegetation status within areas under different management or protection status. This will allow the determination of management profiles that are most successful in protecting ecosystems, including areas best suited for future protection.

The Managed Areas Database (MAD) includes a comprehensive list of the many managed and protected areas in the conterminous United States including National and State Forests, National and States Parks, National Wildlife Refuges, Military and Indian Reservations, and numerous others. Attributes (useful informational data), such as area name, have been entered in MAD for each area. This includes a numeric code for each area which will allow MAD to be automatically linked to the global Protected Areas Data Unit (PADU) database held by the World Conservation Monitoring Centre (WCMC). WCMC was formed by the World Conservation Union (IUCN). The basic informational format of MAD was modeled after WCMC's global database. An additional management-level classification (based on the US

Gap Analysis Project (GAP)) was formed for the areas in MAD.

MAD was created at a coarse intended map scale of 1:2,000,000 with a minimum mapping unit (MMU) of about 1 square kilometer. Areas which were not large enough to meet the MMU were entered into a point data coverage.

Current plans call for a Beta version of the MAD is to be released to knowledgeable reseachers, and, incorporating their feedback, MAD will be updated. It is expected that continuous updates of MAD will be required as boundaries and information change over time.

Remote Sensing Research Unit
1629 Ellison Hall
Department of Geography, University of California
Santa Barbara, CA 93106-4060
Telephone: (805) 893-3845
Fax: (805) 893-3703
Internet: gavin@geog.ucsb.edu
http://www.ncgia.ucsb.edu/rsru/rsbios.html#mcghie

Gretchen G. Moisen , D. Richard Cutler, and Thomas C. Edwards, Jr.

Analyzing Thematic Map Accuracy Using Generalized Linear Mixed Models

---

Abstract:

With the increasing demand for broad-scale vegetation maps for ecosystem management and conservation planning comes the need for flexible tools to assess thematic accuracy of these maps. In this poster, we use a generalized linear mixed model (GLMM) to explore the relationship between thematic accuracy in the blackbrush cover-type of a satellite-based vegetation map of Utah (Homer et al. in press) and various topographical and heterogeneity components of that map. Because of the difficulty in accessing many rugged areas of this State, two strata were defined based on proximity to roads. Vegetation type was recorded on heterogeneous linear clusters of sample points within the "off-road" strata, and on randomly distributed sample points within the "road" strata on selected USGS quadrangle maps (Moisen et al. 1994, Edwards et al. in review). A binary response (correctly classified / incorrectly classified) was modeled as a function of both fixed and random effects accounting for spatially autocorrelated observations and different covariance structures for the random effects. The modeling exercise suggested a strong relationship between map error in the blackbrush cover-type of the Colorado Plateau of Utah, and stratum, slope and local heterogeneity. See Moisen et al. (in press) for further details.

References:

Edwards, T. C., Jr., G. G. Moisen, and D. R. Cutler. (In review) Assessing map uncertainty in ecoregion-scale cover-maps. *Ecological Applications*.

Homer, C. H., Ramsey, R. D., T. C. Edwards, Jr., and A. Falconer. (In press) Landscape cover-type modelling using a multi-scene TM mosaic. *Photogrammetric Engineering and Remote Sensing*.

Moisen, G. G., T. C. Edwards, Jr., and D. R. Cutler. (1994) Spatial sampling to assess classification accuracy of remotely sensed data. Pages 161-178 *in* J. Brunt, S. S. Stafford, and W. K. Michener, eds. *Environmental Information Management and Analysis: Ecosystem to Global Scales*. Taylor and Francis, Limited.

Moisen, G. G., D. R. Cutler, and T. C. Edwards, Jr. (In press) Generalized linear mixed models for analyzing error in a satellite-based vegetation map of Utah. *Proceedings of the Second International Spatial Accuracy Symposium*, Fort Collins, Colorado, May 21-23, 1996.

Authors:

Gretchen G. Moisen , USDA Forest Service, Intermountain Research Station, Ogden, UT 84401 801-265-5384 (PH), gretchen@nr.usu.edu (E-MAIL)

D. Richard Cutler, Department of Mathematics and Statistics, Utah State University, Logan, UT 84322-3900, 801-797-2699 (PH), richard@sunfs.math.usu.edu (E-MAIL)

Thomas C. Edwards, Jr., USDI National Biological Service, Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, UT 84322-5210, 801-797-2529 (PH), tce@nr.usu.edu (E-MAIL)

Igor S. Pashkovsky, Egenea Yu. Potapova, and Boris A. Shmagin

DYNAMIC-STOCHASTIC MODEL OF GROUND WATER BALANCE FOR RIVER BASINS

INTRODUCTION

Spatially-explicit modeling of the ground water balance of river basins forms a basis for environmental monitoring and for estimating the distribution of water resources within river basins. Such information is crucial for planning human activities that affect water use, such as pumping from wells.

Many different hydrogeological models exist with complex mathematical solutions, but most have not been tested with empirical data. We have developed a spatially-explicit mathematical model which uses simple underground flow equations and unsaturated zone equations to calculate all elements of ground water balance, and have verified its performance using detailed, long-term hydrologic records from several Russian river basins. The model consists of five elements: climate (temperature and precipitation), boundary conditions of the river basin surface, and conditions of the unsaturated zone, surface and underground flow. River basins are gridded into square cells for model computation, and input parameters are defined by basin relief and geomorphology. An interpolation subroutine estimates basin parameters where field data are lacking.

Previous versions of the model (Pashkovsky 1988) did not consider unsaturated flow. The current version makes it possible to calculate all elements of ground water, and simulate and estimate the influence of diverse basin characteristics and human activities. A key feature of the model is its treatment of snow cover and frost depth, which greatly influence surface and ground water runoff in regions with cold climates such as Russia.

The model has been packaged with subroutines for input and graphical display: "Software System for Modeling Groundwater Flows and Migration in Saturated/Non-Saturated Rocks" (Burlin et al. 1994). The model generates output values (surface and ground water runoff) for each cell in a gridded river basin, and estimates the river hydrograph at the mouth of the basin. Model output can be shown using the program's display routines, or can be exported as an ASCII file for display and manipulation in a GIS.

Applications

Example applications of this model are shown for the Little Istra and Usadevscy Rivers. The basin of the Little Istra River includes 13 sub- basins which differ in natural, geological, hydrogeological and land cover conditions. We have realized models for ten of these sub-basins, and show the results from two with different erosion and aquifer discharge rates. These two sub-basins, the rivers Kozinca and Kocheiscy, are almost equal in area (6.48 km2 and 8.25 km2, respectively) and in land cover (> 95% forest), but their geology differs. The upper part of the geological column (10-15 m) is clayey ground moraine in the Kozinca River

Basin, and glaciofluvial sand in the Kocheiscy River Basin. Other parameters for the unsaturated zone are similar in the two basins, except for the thickness and permeability coefficient of the undersoil strata, and the connection coefficient with the underlying aquifer. The model shows that ground water fluxes (contribution to river flow, recharge of the upper aquifer, losses to the lower aquifer) are much higher for the Kocheiscy River Basin than the Kozinca River Basin. The model agrees well with natural data.

The Usadevscy River Basin drains into Lake Valdai, and is very marshy (> 30% wetland). We used the model to estimate: (1) ground water contributions to Lake Valdai, and (2) the effect of wetland drainage on river flow. This analysis showed that ground water contributions to river flow are not more than several millimeters per year, equal to only one percent of the annual runoff to the lake. Simulated drainage of the wetlands in the river basin by surface ditches 2 m deep and 40 m long considerably increased ground water fluxes to the river and recharge of lower aquifers. The results confirm the sensitivity of the model to different landscape types and human modifications of the river basin.

REFERENCES

Burlin M.Yu., L.A. Panova, and I.S. Pashkovsky. 1994. The Software System for Modeling Groundwater Flows and Migration in Saturated/ Non- Saturated Rocks. pp.164-176. in: Proceedings of the Workshop on Hydrogeology/Environmental Geology Modeling (August 21-27, 1994), Zdarske Vrchy, Czech Republic.

Pashkovsky, I.S. 1988. Estimated recharge of underground waters and mathematical modeling of surface and underground runoff from river basins. pp. 79-103. in: Methods of waterchange study. Naucova Dumca, Kiev.

AUTHOR INFORMATION

Igor S. Pashkovsky, Dr.Sc.
Professor
Department of Hydrogeology
Moscow State University
Moscow, 119899
RUSSIA
Phone: 095/115-9986
FAX: 095/115-9992

Egenea Yu. Potapova
Scientist
Department of Hydrogeology
Moscow State University
Moscow, 119899
RUSSIA
Phone: 095/939-2112
FAX: 095/939-3108

Boris A. Shmagin, Ph.D.

Senior Scientist
Department of Hydrogeology
Moscow State University
Moscow, 119899
RUSSIA
Phone: 095/203-2936
FAX: 095/203-6578
Email: shmag@hydro.geol.msu.su

Micha Pazner and Brian Reynolds

The "Clear Box" Image Processing Simulator: A Physical Device for Demonstrating Digital Image Processing Functions

---

Clear Box is a prototype physical device we designed and built for use in the classroom to demonstrate digital image processing functions. We are using the device to teach remote sensing, GIS, and cartographic image processing operations.

The image shows Clear Box sitting on top of an overhead projector. The box is made of transparent material and has three 'layers' - sliding trays with labeled tabs. Each layer has place holders for 16 'cells' - 35mm slide frames that have vertically-dis placed color-coded numbers on them. The slides are kept, organized by value and by layer, in an accompanying slide storage box. The device lends itself to interactive demonstration of digital image processing transformations involving up to three layers .

The demonstrated examples are planned in advance. Initial examples involve basic operations, while later examples take advantage of the device to illuminate confounding operations such as neighborhood filters. Clear Box is used when appropriate rather t han in every lecture, and is integrated with other classroom media such as blackboard, overhead transparencies, and computer overhead projection.

The image illustrates the use of Clear Box in the classroom. There are several advantages to using the device: it makes the abstract model of a stack of layered cellular arrays tangible, it focuses attention on the micro level (cell, cell neighborhood), and stresses numeric values as the primary data representation (rather than colors). Clear Box helps take some of the 'black box' aspect out of image processing.

What is the connection between Clear Box and environmental modeling? Image processing operations are the building blocks of cellular environmental models, and Clear Box is an instructional tool for learning such operations. The device is being used to augment other teaching aids, including a user-oriented operation taxonomy, along with associated modeling exercises. Clear Box may help budding environmental modelers on their way to becoming spatial problem solving maestros.

Micha Pazner (Associate Professor) and **Brian Reynolds** (Student), Department of Geography, Social Science Centre, The University of Western Ontario, London, ON N6A 5C2 tel: 519 661 2111 x5031, fax: 519 661 3750 email: pazner@sscl.uwo.ca

Eva-Maria Stephan

Visually Exploring Mutual Relationships in Environmental Data

Abstract

Environmental sciences include the study and modeling of spatial and temporal natural phenomena on the earth surface. Model building requires the integration and analysis of various phenomena, which in GIS usually are represented as sampled or simulated data fields, e.g., on a regular or non-regular grid, or as point data. While GIS are often used for integrating heterogeneous data, they still lack on the facility to *visually explore relationships* between multivariate data.

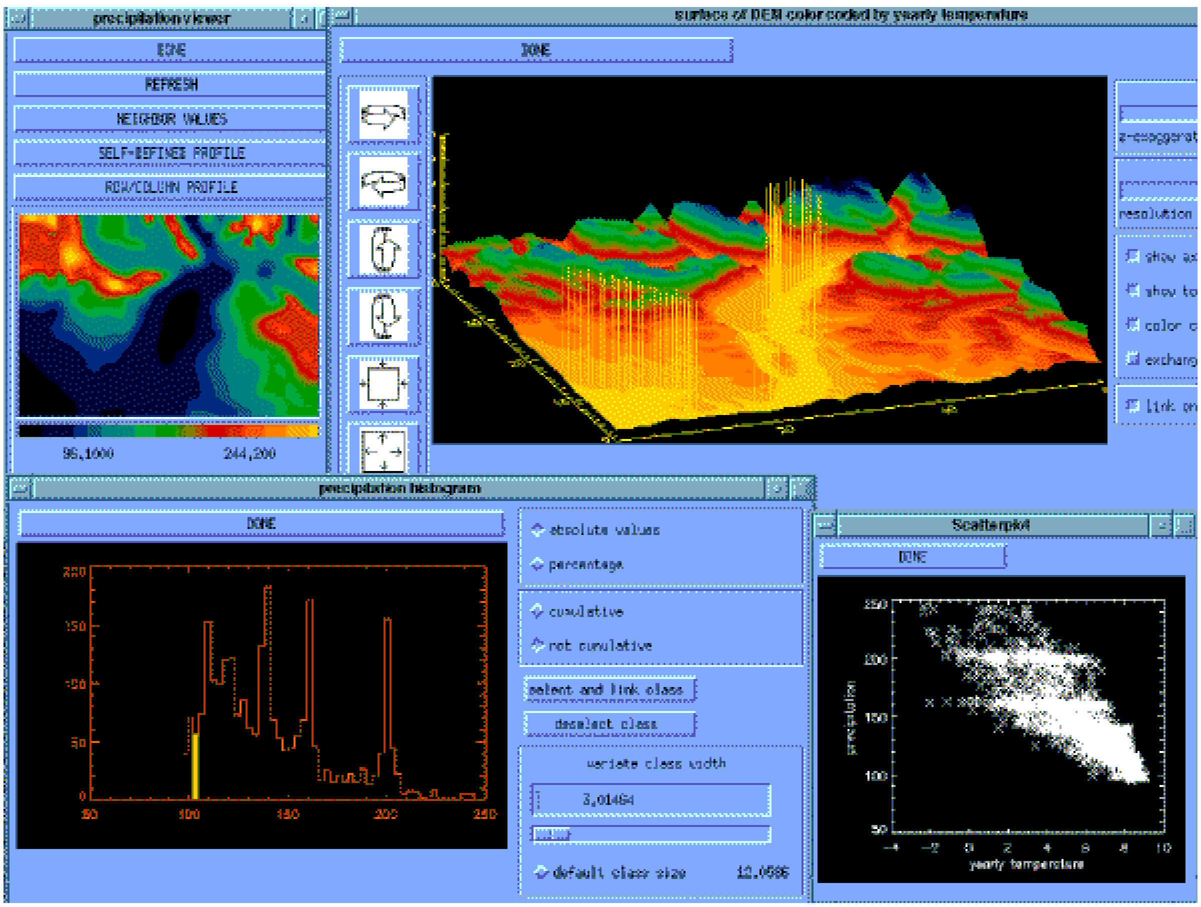
Typical visualization methods in GIS are restricted to static 2- and 3-dimensional base maps which may be overlaid by vector graphics; visualization mainly serves descriptive and presentation purposes, where the phenomena represented in the data are already known, and are communicated clearly to others through a visual presentation. Beyond that, GIS do not support the use of graphics for viewing and investigating several continuous and/or contiguous data layers simultaneously. There is a great need to make dynamic graphics part of the (data) analysis process and support *Exploratory Visualization (EVIS)*. EVis in GIS requires to integrate concepts of direct manipulation, data flow and multi-dimensional visualization techniques for spatio-temporal data. Such techniques can be derived from the field of scientific data visualization but must be further developed for environmental data analysis.

The poster (get poster A, poster B) describes graphical techniques for *exploratory visualization of environmental data* and introduces a prototype system, called *DataScaping*. DataScaping is an interactive visualization system for viewing and for exploring mutual relationships in environmental data. Surface data is presented graphically, but other than than photorealistic landscapes *"all kinds of data-surfaces"* can be generated, browsed and manipulated interactively. The introduced graphical toolbox includes data visualization by 3-d surface rendering and animation, as well as statistical plots. Main emphasis is put on the *interactive manipulation* of views through a graphical user interface (GUI), and the *dynamic linking* of data in multiple windows. Requirements for interactive manipulation include the change of viewing parameters (perspective, position of object, animation time) and object (exaggeration, color, certainty level, resolution, data values). The dynamic linking, based on *dynamic data referencing*, provides a link between various representations, e.g., statistical images and pseudo-realistic surface views.

In the presented approach of exploratory visualization for environmental data, graphics are used for directed or indirected search, when the data analyst does not (always) know what he is looking for, and visualization helps him to find patterns and relations in the data and to gain

understanding and insight in to the overall nature and relationships of multivariate data. The presentation in gridded or rendered (and animated) surfaces provides an intuitive metaphor for exploring spatio-temporal data. In addition, interaction and selection tools are developed, which are necessary to visually explore data with higher dimensionalities, i.e., multivariate data sets.

Exploratory visualization with the Datascaping prototype system is a promising new spatial data analysis tool. Presented techniques for the investigation of environmental surface data offer potential for better exploring interdependencies and characteristics of multivariate data, by providing the possibility to intuitively browse multiple heterogeneous data layers simultaneously. Various examples, such as the interactive exploration of a temporal glacier data set, or exploring two differently interpolated temperature data sets are given in the poster (get poster A and poster B). The image below shows a screenshot of a typical DataScaping session.



The image shows a screenshot of a typical DataScaping session. Relationsships between a precipitation, a mean temperature and a DEM data set can be explored in multiple linked windows (data resolution: 1 km raster). The geographical location of the selected values in the precipitation histogram plot (lower left) are highlighted in the color-coded temperature surface (DEM with draped temperature map, upper right). The upper left image shows the precipitation data set as color image with the posssibility to interactively query neigbor values and generate cuts and draw profiles. The lower right is a scatterplot of the two attributes.

Eva-Maria Stephan
Spatial Data Handling
Department of Geography
University of Zurich
Winterthurerstr. 190, 8057 Zurich, Switzerland
tel: +41 -1 -257 5255, fax: +41 -1 -362 5227
email: stephan@geo.unizh.ch

David M. Stoms, Frank W. Davis, Richard L. Church, William. J. Okin, and Joe Walsh

Selecting Biodiversity Management Areas for the Sierra Nevada Region

The objective of this analysis was to explore opportunities for siting new Biodiversity Management Areas (BMAs) in the Sierra Nevada region. We define BMAs as specially designated public or private lands with an active ecosystem management plan in operation whose primary purpose is to contribute to regional maintenance of native genetic, species and community levels of biodiversity, and the processes that maintain that biodiversity. The strategic goal was to design a BMA system that represents all major Sierra plant community types at some predefined fraction of their mapped distribution. The approach involved a loose coupling of a GIS with an optimization program we call the Biodiversity Management Area Selection (BMAS) model.

We present a planning process comprised of five basic steps. The first three steps involve data preparation, the fourth is the BMAS model, while the fifth step is the evaluation of results. In Step 1, the vulnerability of every biodiversity element is determined by a GIS overlay of maps of biodiversity distribution and existing BMAs, and thus constitutes the assessment phase of a gap analysis. A target level for representing biodiversity elements is selected. An element is termed *vulnerable* if the area of its distribution within areas formally managed for long-term protection is less than this threshold. *Remaining vulnerability*, therefore, is the difference between the desired representation for each vulnerable element and its current representation. Step 2 involves the disaggregation of the unprotected area of each vulnerable element to the planning units in which it occurs. This disaggregation identifies the opportunities provided by each planning unit to reduce the remaining vulnerability. Whereas steps 1 and 2 provide the biological information needed for conservation planning, in step 3 we introduce the factors that affect the suitability of planning units for biodiversity management. This suitability of a planning unit depends on a variety of factors that characterize it such as who owns it, how impacted is it by humans, and how manageable is it for biodiversity? Steps 1-3 are performed within the GIS. The BMAS model is executed as Step 4 of the process. Because we might consider hundreds of elements to be vulnerable and can select from among hundreds of planning units, the problem is relatively complex. Although the model can be implemented as an integer programming (IP) model, the current version uses a heuristic approach. We then use this optimization model to select additional planning units (small watersheds in this case) to achieve full representation at the desired level as efficiently as possible. That is, the total area allocated to BMAs is minimized while also creating the least conflict with roads, people, and private land ownership. Information about selected watersheds is passed back to the GIS for display and further analysis. A wide range of alternative BMA systems were examined based on different assumptions, constraints, and target levels for representation. Here we show the results for one such BMA system for the northern Sierra Nevada subregion.

Acknowledgments

Financial support for development of the BMAS model was provided by the U. S. Forest Service Sierra Nevada Ecosystem Project. Computing facilities were provided by a grant from the IBM Corporation Environmental Research Program. Biodiversity and land management data were developed for the California Gap Analysis Project at UCSB and coordinated by the National Biological Service. Suitability data were compiled as part of a cooperative agreement from the Environmental Protection Agency. Thanks to the many research assistants in the UCSB Biogeography Lab who helped compile these datasets. We benefited greatly from advice from the SNEP science team, especially Dr. Norm Johnson of Oregon State University.

Author Information

David M. Stoms, Assistant Research Scientist
Institute for Computational Earth System Science
University of California, Santa Barbara
Santa Barbara, CA 93106-3060
Voice: (805) 893-7655
Fax: (805)-893-3146
e-mail: stoms@geog.ucsb.edu

Frank W. Davis, Professor
Department of Geography
University of California, Santa Barbara
Santa Barbara, CA 93106-4060
Voice: (805) 893-3438
Fax: (805)-893-3146
e-mail: fd@geog.ucsb.edu

Richard L. Church, Professor
Department of Geography
University of California, Santa Barbara
Santa Barbara, CA 93106-4060
Voice: (805) 893-4217
Fax: (805)-893-3146
e-mail: church@geog.ucsb.edu

William J. Okin, Graduate Student Researcher
Department of Geography
University of California, Santa Barbara
Santa Barbara, CA 93106-4060
e-mail: okin@geog.ucsb.edu

Joe Walsh, Staff Research Assistant
Institute for Computational Earth System Science
University of California, Santa Barbara
Santa Barbara, CA 93106-3060

Voice: (805) 893-7044
e-mail: walsh@geog.ucsb.edu

D.A.Swayne, D.C.L Lam, J.D. MacNeil, Adrian Harding, Mark Mayo, Alex Storey, Ken Brown, Doug Kay

INTEGRATING NOVEL APPLICATIONS INTO A GIS FRAMEWORK: SOME EXAMPLES

ABSTRACT

We demonstrate two examples of novel software tools which are relevant to GIS / modelling applications.

An air trajectory is the pathway that a parcel of air takes as it passes through the atmosphere. An air trajectory has both horizontal and vertical motion in the atmosphere. Trajectories are large (up to 84-parameter space for 4-5 day, 6 hour sampling), and numerous. They represent a real challenge in proposing and implementing a classification algorithm which can provide useful information in reasonable time. A number of techniques are looked at to cluster the air trajectories, including hierarchaical, k-means and LVQ. The information a clustered set of trajectories can provide is a possible source for another set of data. For example, by comparing ozone data collected at a site with the clustered trajectories, a possible source for the high ozone concentrations could be located.

Such analysis can form the backbone of the conversion between fixed grid (Eulerian) methods for solving transport equations in geographical space - time and moving (Lagrangian) methods which characterize the sources and receivers of air pollution. Methods for effecting this analysis can have a far-reaching effect on the understanding of the means and the efficacy of air pollution reduction.

Included in this poster will be a software system, for Windows on Personal Computers, which can be used for the plotting and visualization of atmospheric trajectories which have been calculated from wind vector files obtained by weather observation and modelling. This software can have an important role in determining souces, transport and fate of airborne contaminants. It can also be viewed by accessing

http://cfc.crle.uoguelph.ca/

Earlier versions of it are available for trajectory analysis, and new functionality is developed on a regular basis. It comes with its own rudimentary GIS and can be adapted to other software systems.

A second example of a GIS-ready application is our "Manure Wizard", a decision support prototype for a system based on expert knowledge of management of farm waste, optimal use of this waste for crop fertilizer, and assessment of its impact on the environment. This system is based on extensive work of agricultural extension specialists (most notably Agriculture

Canada and Ontario Ministry of Agriculture, Food and Rural Affairs).

Its potential for modelling and assessment and minimizing the impact of nonpoint source pollution is tremendous, since a record of its use can be a highly accurate data source for farm chemicals, crop and waste distribution (with the cooperation of collaborating farmers and other end users). Also, it has the potential ability to give precise recall of relevant information to the farmer, with the latest research results available online.

D.A. Swayne , University of Guelph, Guelph N1G 2W1, CANADA
dswayne@uoguelph.ca

D.C.L Lam, National Water Research Institute, Environment Canada, Burlington L7R 4A6 CANADA
David.Lam@cciw.ca

J.D. MacNeil, University of Guelph
jd@uoguelph.ca

Adrian Harding, University of Guelph
adrian@quickweb.com

Mark Mayo, University of Guelph mark@ozone.crle.uoguelph.ca

Alex Storey, ES Aquatic Inc., Kitchener, Ontario, CANADA

Ken Brown, ES Aquatic Inc., Kitchener, Ontario, CANADA

Doug Kay, National Water Research Institute, Burlington, Ontario, CANADA

MERGING TWO LARGE MINERAL LOCATION DATABASES USING LOGISTIC REGRESSION

Deborah J. Shields

U.S.D.A. Forest Service, Rocky Mountain Forest and Range

Experiment Station, Fort Collins, CO, 80526

Stella W. Todd

Management Assistance Corp. of America

3825 E. Mulberry St.

Ft. Collins, CO 80524

Douglas D. Brown

U.S.D.A. Forest Service, Rocky Mountain Forest and Range

Experiment Station, Fort Collins, CO, 80526

ABSTRACT

The purpose of this project is to merge two large mineral information databases: the Mineral Resource Data System (MRDS) from the U.S.D.I. Geologic Survey (69,502 records), and the Mineral Availability System/Mineral Industry Location System (MAS/MILS) from the U.S.D.I. Bureau of Mines (208,065 records). Each database contains unique mineral location data; however, some sites are present in both databases. Prior to merging the databases, it is essential that we identify all those mineral locations occurring in both databases so that only unique information is contained within the final single database. Although the full extent of overlap between the two databases is unknown, the two agencies have cross-referenced 21,344 duplicate mineral records in the 1994 MAS/MILS and 1993 MRDS databases. Logistic regression has been selected as the classification approach with which we will determine whether pairs of mineral location records, one selected from each database, in fact represent the same site. The logistic model follows:

$\log(i / 1 - i) = \mathrm{b}k\ ik\ \ i$

where: i = probability value

b*k* = slope coefficient

ik = explanatory variable

This statistical method estimates the probability of there being a discrete response to a set of explanatory variables, which themselves can be binary or continuous. The formula for calculating the probability is: *i* = exp(*i*) / (1 + exp(*i*). In our application, outcome classes were binary: matched or same site (value = 1) versus unmatched or not the same site (value = 0).

Possible explanatory variables were mineral location attributes common to both databases, a set which includes name of the site, mineral commodities present or extracted, state, and county, as well as site location (x and y coordinates). Some of the explanatory mineral attributes were binary, for example, mineral commodities which were either the same (value = 1) or different (value = 0).

The frequency distributions of 5000 known matched and 5000 known unmatched sites were compared to determine if a categorical distance variable would be a useful discriminator of matched sites. Because all possible unmatched site pairs were sampled within a 1 km window, we observed that the frequency distribution of between-pair distances (0-1 km with .1 km intervals) was relatively flat. In contrast, the frequency distribution of matched site pairs was skewed toward small between-pair distances, with about half of the matched mine pairs occurring within .1 km of each other. This distance was selected as the threshold value for determining a distance match. Separation distances <= .1 km were assigned a distance match code (DMC) value of 1 (one). The non-match value was 0 (zero).

Site names were shortened and concatenated using "C" programs. The purpose of the shortening was to remove some of the between-database variability associated with spelling and nomenclature. Single symbol diphthongs were substituted for two letter groups to speed up processing time. Site names were compared for the percentage of identical letters and symbols within each string. The short site name percentage (SNP) variable

reflected this value.

After removing internal duplicates, five thousand (5000) known matched pairs of mineral sites were randomly sampled from within the 21,344 cross-referenced records. Pairs were composed of one MRDS and one MAS/MILS site. In contrast, unmatched pairs were composed of either two MAS or two MRDS sites, again after removal of internal duplicates. Five thousand (5000) known unmatched pairs of mineral sites were randomly sampled

proportionately from MAS/MILS and MRDS. They were selected in Arc/Info, using a search window within 1 km of x and y coordinate locations.

A logistic regression function was determined from the 5000 matched and 5000 unmatched sets of mineral attributes, using a stepwise procedure. The percent matches for both the shortened and original site name strings were tested in the development of

the logistic function. The shortened names were selected in the stepwise logistic process, and were more highly correlated with matches than the long original names (R = 0.73 and R =

0.56 respectively). Of the three selected variables two were binary: commodity name match code (CNMC), and distance match code (DMC), and one was continuous: short site name percentage (SNP). The following logistic regression function resulted: $i = -4.4297 + 0.5555(CNMC) + 5.3538(SNP) + 2.3510(DMC)$, (Pr > Chi- square 0.0001 or less).

The attributes of each MRDS mineral location will be compared with those MAS/MILS locations falling within a 1 km search window. The explanatory variables generated by this

comparison will be fed into the previously developed logistic regression function. Those MAS/MILS records with the highest probability of matching a MRDS record will be identified and the probabilities recorded in a separate attribute field. These

probability values will be used as the criteria for merging the two databases.

ACKNOWLEDGEMENTS

Portions of this research project were conducted by Dr. Denis Dean and Mr. Brian Peters under Cooperative Agreement (No. 28-C3-764) between Colorado State University, Department of Forest Sciences, and the USDA Forest Service Rocky Mountain Forest and Range Experiment Station Research Work Unit No. 4852. Portions of the research reported here was funded by the USDA Forest Service Rocky Mountain Forest and Range Experiment Station at Fort Collins, Colorado under contract with Stella W. Todd through the Management Assistance Corporation of America. This research is being conducted as part of the Minerals Technical Document supporting the USDA FS Resource Planning Act Assessment.

Fill in your name and email address here

GIS-Based Modeling of Desert Tortoise Habitat in the Mojave Desert

Joseph M. Watts

Natural resource managers responsible for endangered species management can benefit from GIS-based wildlife habitat models. On that premise, a geospatial habitat model of a Desert Tortoise population on the National Training Center (NTC) at Fort Irwin is under construction. The model's function is to statistically relate field mark-and-recapture tortoise data to GIS layers on plant comunities, soils, topography, and geology. The primary data sources for the project are remotely sensed imagery, GPS survey data, and field transect data. Numerous organizations are participating in the project including the U.S. Army Topographic Engineering Center, National Park Service, Natural Resources Conservation Service, NTC, and members of the academic community.

Visual representation and analysis of the climatic data using GIS

Demonstration

Harumi Kitajima Yanagimachi, Shinshu University
Kazutaka Iwasaki, Hokkaido University
Kenji Sato, Pasco Corporation

Abstract

The purpose of this study is to show our climatic map databases of Japan and the animation of the 2-D or 3-D climatic maps and point out the advantages and disadvantages of the application of GIS to the climatic change researches. We have been constructing climatic databases of Japan and studying the effective visualization methods of them to be used for the climatic studies. These databases comprise many digital maps of climatic elements and are managed by Arc/Info.

We pick up snow depth data and investigate the day-to-day distribution changes and year-to-year variations. Using 2-D or 3-D map animation, these changes or variations are easily understood and then it is concluded that these methods can be applied to studies of climatic or environmental changes.

There is a remarkable difference of surface condition in winter between the Pacific side and the Japan Sea side in Japan, because the main dividing ranges are lying from north to south in Honshu (the main island of Japan) and these barriers prevent the winter monsoon westerly flow. Therefore there is much snowfall on the west of the dividing ranges, while it is clear almost everyday on the east of them in winter. This contrast is influenced not only by the mesoscale circulation but also by the global circulation.

The distribution maps of snow cover extension and snow depth are efficient for climatological and hydrological studies. It is especially useful for the predictions of the global circulation. The climatic modeling considered by snow cover extension will contribute to the prediction for the future climatic change including greenhouse-effect warming. We are also analyzing the relationship between the snowfall and the other elements such as temperature and precipitation.

To grasp the distribution of the climatic elements instantly is effective for the climatic change research. In addition to the 2- D and 3-D maps, animation movies are recently attracted as the better representation to understand the various phenomena, because animation is one of the methods that include the time dimension. We are making the mpeg movies of snow depth distribution, too. These movies are created from the images made by Arc/Info and displayed on the PCs or workstations using public domain software or freeware. We will introduce how 3-D and animation presentations have advantages in the fields of climatic studies, too.

1. Visualization of the Snow Depth Distribution in Japan using Arc/Info

Harumi Kitajima Yanagimachi & Kenji Sato

- Snow cover climatology and GIS (3.8MB, PowerPoint ver.4.0)

QuickTime movies of the daily snow depth distribution in Japan

- 1989/1990 winter (5.4MB)
- 1990/1991 winter (5.3MB)
- rotating view (3.1MB)
- December 15, 1954-1990 (1.5MB)
- January 15, 1955-1991 (1.5MB)
- February 15, 1955-1991 (1.5MB)
- March 15, 1955-1991 (1.5MB)

2. A Case Study on the Distribution of Snow Depth in Hokkaido

Yuichi Hashimoto & Kazutaka Iwasaki

- QuickTime movie (17.9MB, Macintosh only)

Harumi Kitajima Yanagimachi
Faculty of Economics, Shinshu University
3-1-1 Asahi, Matsumoto, Nagano 390 Japan
Phone: +81-263-35-4600 ex.3335
Fax: +81-263-36-7220
e-mail: yanagi@econ.shinshu-u.ac.jp

Kazutaka Iwasaki
Faculty of Information, Shizuoka University
836 Ohya, Shizuoka-shi, 422 Japan
Phone and Fax: +81-54-238-5618
e-mail: iwasaki@ia.inf.shizuoka.ac.jp

Kenji Sato
Pasco Corporation
Higashiyama Building
1-1-2 Higashiyama, Meguro-ku, Tokyo 153 Japan
Phone: +81-3-3715-1221
Fax: +81-3-3715-1421
e-mail: st2@pasco.co.jp

CREATION OF A 3D PERSPECTIVE CLASSIFIED FOREST MAP USING GEOGRAPHIC INFORMATION AND REMOTE SENSING INTEGRATION

H. Yildirim, E. Alparslan, B. Bilge, H. Kurar, O. Divan, S. Elitas

TUBITAK-MAM Space Technologies Department
P.O. Box 21 41470 Gebze-Kocaeli / TURKEY

ABSTRACT

New possibilities for forest mapping using integration of geographic information systems (GIS) and remote sensing (RS) technologies are investigated. RS data processed by image classification methods is subsequently converted into a GIS data coverage, offering the possibility of analyzing the classified image together with other data levels such as digital terrain model (DTM) data existing in GIS. It was possible to analyze each forest type according to the altitude information in GIS, i.e. certain tree types would only grow in certain altitudes. As a result, the information extracted was tabulated.

The study area is located on Black Sea Shore, Turkey, being a small part of Ordu province covering 20 km by 22 km area with 40º 50' N / 41º 10' N latitudes and 37º 44' E and 37º 58' E longitudes. LANDSAT TM satellite imagery corresponding to the study area, acquired on 15.9.1987 was classified using ground truth information collected from Directorate of Forest Works (DFW) of that area. The classified satellite image was transferred into GIS as a data layer. Then, topographic map sheets of the study area obtained from General Commander of Mapping (GCM) at 1:25.000 scale were digitized at 50 m intervals to extract contours of constant elevation of the study area. Land cover was draped over the surface model generated from contours of constant elevation, creating a 3D perspective image of the area to facilitate further analysis.

Known UTM coordinates collected from map sheets of GCM were used as ground control points in geometrical correction of the satellite image corresponding to the study area in order to obtain an image georeferenced into its proper UTM coordinates. Maximum likelihood classifier was trained by ground truth information collected from DFW of the area and the training samples were validated using a minimum distance classifier. Validation served as some filter excluding the outliers in the training areas. The classifier assumed that each class would have normal (Gaussian) distribution and assigned each pixel to a class having the minimum Mahalanobis distance. Seven different land covers identified in the study area were: sea, alder forest, chestnut - alder, healthy mixed forest, unhealthy mixed forest, urban area and farm land.

The classified image was visually analyzed and some tiny class polygons in the form of pixel clusters were discovered with a center pixel being connected to neighboring pixels in horizontal, vertical and diagonal directions. Since these polygons are surrounded by larger polygons of some other class, eliminating these polygons will yield a smooth surface in the classified image. A sieve filter was used for this purpose. Due to formation of large class regions by sieve filtering, unit boundary vectors become only visible between pixels assigned to different classes .

The land cover map which is in raster data form was converted into vector form and integrated into ARC/INFO polygon coverage. Arcs forming the polygon coverage were analyzed and arcs sharing pseudo nodes were combined by removing these nodes. The distance between adjacent vertices on splined arc and curves within an added arc is set to 42.42 which is the hypotenus of a right triangle having sides of 30 m, which is the pixel size for TM images. Using this tolerance of 42.42 the shape of arcs building the polygon coverage were changed by building curves at the angles in an arc. Arcs were redrawn so that adjacent vertices are spaced by 42.42 tolerance apart. Vertices were added or deleted depending on this tolerance resulting in arc smoothing or arc generalization, respectively. Smoothed arcs were converted to polygon coverage, resulting polygons having no stair effects on their sides. This coverage enabled further analysis of classified image together with other additional data layers such as digital terrain model (DTM) data existing in GIS.

A surface model was generated from elevation contour lines digitized from topographic maps at 50 m intervals by creating a triangulated irregular network (TIN), which is a series of connected triangles accurately representing a surface with less data points than other data models. TIN creation errors were analyzed generating a descriptive listing about the TIN surface followed by its graphic display. Invalid flat triangles occurring along streams and ridges were eliminated adding new intermediate points along the ridges and streams between the input contours. Additional sample points were entered between the contours in order to increase the distance between vertices on each contour arc, resulting in removal of invalid flat triangles. Weed tolerance and proximal tolerance were also adjusted to remove excess vertices forming flat triangles. The created TIN surface was transformed back to contour arc coverage

Spatial Exposure Analysis of Simulated Oil Spills Using GIS

Miguel Acevedo, Toar Schell, Fred Bogs and Bruce Hunter.
Institute of Applied Sciences and Department of Geography, University of North Texas, Denton Texas, 76203.

A spatial exposure analysis program was designed to quantify, display and compare the impacts of potential oil spills on estuarine ecosystems. The program integrated GIS and the output of an oil spill transport and fate model to summarize temporal contam inant exposure over geographical space. Two approaches were developed and compared: one based entirely on ARC/INFO (ESRI, 1994) macro commands and the other based on a combination of a program written in C language and ARC/INFO commands. In the first a pproach simulated oil spill output from the transport and fate model was converted and processed, completely within ARC/INFO, to calculate spatially explicit exposure over time. The results were then used for overlay analysis with offshore resource cover ages. The C program developed for the second approach was used to save computation time during intensive calculations of exposure. Both vector and grid representations were used in this development. Tampa Bay, Florida, is used as a case study.

Modelling river channel change: lateral instability and risk assessment

Steve Carver1, David Sear2 and Damian Exworthy1

1 School of Geography, University of Leeds, Leeds, UK, LS2 9JT
Tel: ++44 113 2333318, Fax: ++44 113 2333308, Email: steve@geog.leeds.ac.uk

2 Department of Geography, University of Southampton, Southampton, UK, SO9 5NH
Tel: ++44 1703 594614, Fax: ++44 1703 593729, Email: d.sear@soton.ac.uk

Abstract

This paper outlines a GIS approach to modelling laterla instability in river channels as a means of risk assessment in flood plain management. The work of Graf (1981, 1984) and Hooke and Redmond (1989) in modelling lateral instability in gravel bed rivers is adapted to produce risk maps through the incorporation of impedance models and land value information. Graf's raster based probabilistic approach to predicting instability on the basis of historical data derived from old maps and aerial photographs is adapted and vector alternatives investigated. Improvements to model predictions are made through the development of basic rule based impedance models describing the effect of changes of bank material, riperian vegetation, topography and river engineering on channel movement and instability. Results from the combined models are integrated with land use maps tagged with land value information in order to estimate risk in relative monetary terms. test case examples from the River Wooler and River Ure in northern England are presented and used to evaluate the potential of the technique as a tool for flood plain management.

Raymond J. O'Connor and Malcolm T. Jones

Scale and the ecologically-relevant characterization of landscape pattern.

Abstract:

Why have there been so few empirical demonstrations of relationships between animal distributions and the metrics popularly used to characterize landscape patterns? One would expect many, given a) how often spatially explicit models demonstrate crucial contribution from spatial patterning of resources, and b) the widespread advocacy of metrics such as dominance, contagion, and fractal dimension as characterizations of natural landscapes. We used classification and regression tree (CART) models of the distributions of some 600 bird species across the coterminous United States to demonstrate that environmental correlates of different types manifest themselves only with characteristic scales. We present hierarchical models that show how HCN climate data and anthropogenic stressor data and remotely sensed land-use and landscape pattern data interact in their constraints. The models differ from species to species but show recurrent patterns of constraint that constrain bird distributions within homogenous regions. These regions do not universally coincide with established ecoregion classifications. We conclude by discussing the implications for future analyses of the geographical distributions of animals.

D.C.L Lam, D.A. Swayne, C.I. Mayfield

Environmental Data Sharing - Software and Protocols

MapMosaic is a network-aware program that acts as a catalog for retrieving data and as a browser of georeferenced information and data. It consists of an interface to an SQL database that stores all catalogs, maps, text, images, etc. associated with objects on the maps, and permits local or remote searching and retrieval. It also interfaces to LivePage and LivePage Webmaster (by Information Atrium Ltd.); programs that are also SQL-based. They are used to store all types of information for remote or local SQL inquiry (LivePage) and to import, organize and publish SGML and HTML documents and data on the WWW in SQL format (LivePage Webmaster). Demonstrations of the software and their integration will be presented.

C.I. Mayfield, Professor, Environmental Biology,
University of Waterloo, Waterloo, N2L 3G1
CANADA

D.A. Swayne , Professor of Computer Science, University of Guelph, Guelph N1G 2w1, CANADA dswayne@uoguelph.ca

D.C.L Lam, Senior Scientist, New Technologies Research Branch, National Water Research Institute, Burlington L7R 4A6 CANADA David.Lam@cciw.ca

Object-Orientation in GIS

Description:

Existing GIS' data models are mostly based on the relational data modeling approach. The relational data model does not support all the requirements of GIS as geogrpahic reality cannot be captured efficiently with relational data modeling. This deficiency in capturing geographic reality imposes problems in spatial data manipulation, spatial data analysis, and spatial data processing. An object-oriented data model is more suitable as geographic entities can be represented more naturally. It will allow for a better sematic representation of spatial and attribute data in GIS.

Object-orientation in GIS will also facilitate simpler mechanisms for software interface, software modification, and software extension.

Attendees in this workshop will be introduced to the general concepts and principles of object-oriented paradigm and practical considerations in implementing object-oriented GIS.

Who should attend?

GIS managers/analysts who have a general understanding of existing GIS and would like to learn new technologies and employ them in their day-to-day GIS operations.

Prerequisite: General understanding of today's GIS and their working.

Outline:

- limitations of existing GIS
- introduction to OO paradigm
- OO data modeling
- OO for other aspects of GIS design and development
- examples

Duration: 1 day

Presentors:

Dr. Hassan Karimi, Research Scientist
MCNC- North Carolina Supecomputing Center
P.O. Box 12889
3021 Cornwallis Road
Research Triangle Park, NC 27709-2889
Tel: 919 248-9249
Fax: 919 248-9245
E-mail: karimi@mcnc.org

and

Dr. Y.C. Lee, Associate Professor
Dept. of Geodesy and Geomatics Engineering
University of New Brunswick
Fredericton, N.B. Canada E3B 5A3
Tel: 506 453-5148
Fax: 506 453-4943
E-mail: yclee@unb.ca

SPATIAL ERROR OF POINT DATA: ASSESSING DATA QUALITY

Adam Lewis

The planimetric error of point data is perhaps the simplest form of spatial error to model and analyse, however even this remains an important issue in spatial database applications. The planimetric errors in a set of vegetation quadrat data (155 quadrats) are estimated and modelled with a normal distribution. I demonstrate that the absolute measure of planimetric error is insufficient information to determine the suitability of the point data for different applications - the data quality. However the data quality can be estimated by consideration of the variogram of the data surfaces with which the point data are to be combined. The expected correlation (R-square) of the erroneous values with the true values (at each vegetation quadrat) is primarily determined by the variogram of the data surface, rather than the error model itself, and can be accurately predicted. By various arguments, it can also be shown that there is a close conceptual and mathematical relationship between spatial error and scale. The methodology used provides a simple means for estimation of the quality of point data for modelling applications. The results have important implications for the development and application of standards for spatial data; an absolute statement of error is not an indication of the quality of the data for a particular application.

SCALE-EXPLICIT SPATIAL MODELLING OF ENVIRONMENTAL VARIABLES

Adam Lewis

Environmental varibles can be defined at a range of scales. According to theory, this 'defining scale' may influence the suitability of the variable for prediction, leading to the concept of scale-explicit modelling. The defining scale of several envrionmental variables derived from an elevation model, including elevation, slope, and insolation, is manipulated by filtering of the elevation model using a Gaussian weighted filter. Other variables, such as roughness and topographic wetness index, are manipulated in similar ways. To test the hypothesis that scale-explicit modelling can yield improved results, a simple univariate exploratory data analysis (EDA) is used to determine the scale at which each variable should be defined in order to maximise the accuracy of prediction of abundance of each species of flora.

Following the EDA, decision tree classification is used to derive default-scale and scale-explicit models for species abundance. It is demonstrated that adoption of a scale-explicit approach to modelling can significantly improve model structure and robustness. This is contrary to the a-priori position that finer spatial scale will yield better results. Furthermore, very simple EDA is all that is required to make substantial gains in the modelling phase.

The Integration of Heterogenous Data for Modelling and Analysis

Steve Carty and Barbara Skelly: IBM

The complex analysis of relationships within environmental systems and models often requires scientists to synthesize many disparate types of data. At a minimum, they wish to see their model results in a geographic context. The constant variety of model input & output formats challenges the most adaptable of systems.

This paper presents some issues encountered in case studies. The case study environments include data from groundwater models, weather models, fire behavior models, GIS, satellite images, and USGS DEM's. The issues involved with integrating various data types to model, analyze, and visualize environmental systems include:

common data formats
unique data formats
integration effort
degree of integration

The paper also presents a suite of data conversion utilities, many developed as a result of these studies. The utilities convert various common data formats into Data Explorer (DX) format. DX is a scientific visualization tool. It's object-oriented data structures and extensive n-dimensional analysis, display, & query capabilities make it a powerful integration target.

The fate of GRASS GIS as a Public Domain Technology Test Bed and Enhanced GIS.

David Hastings NOAA National Geophysical Data Center James Westerveldt, Univerisity of Illinois/US Army Corps of Engineers- CERL

Subtitle: Can a Web-based distributed management approach (similar to Linux's) be GRASS' future driving force?

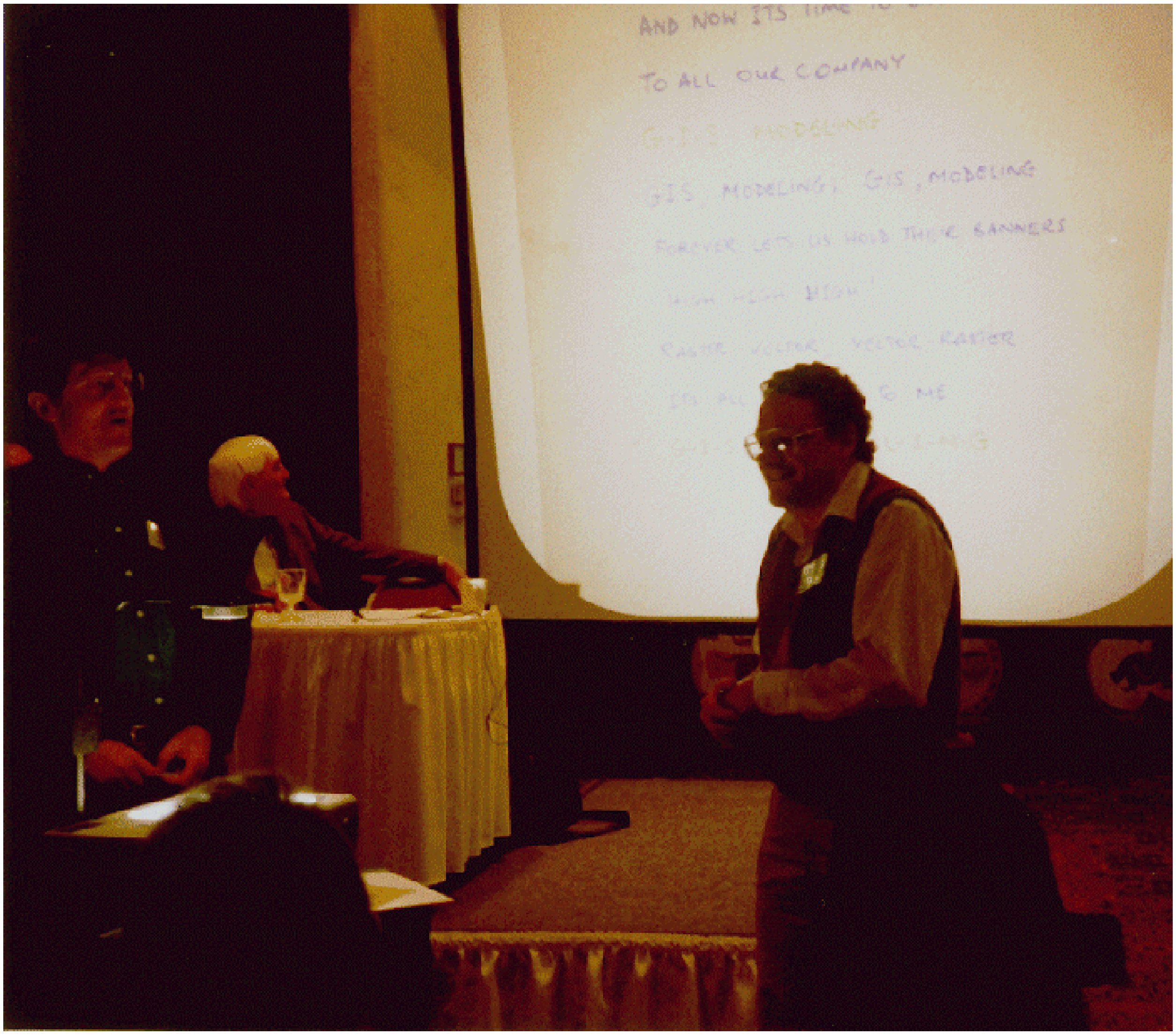
Abstract:

The Geographic Resources Analysis Support System (GRASS) GIS has been under development for 14 years by the Construction Engineering Research Laboratories (CERL) of the Army Corps of Engineers. But CERL's mission is evolving as is its commitment to GRASS. While internal development efforts may proceed at CERL, it will not continue to develop a public domain GRASS. Instead CERL will transition users to commercial off-the-shelf (COTS) software products or to a commercial GRASS yet to be realized through private enterprise.

GRASS was created after CERL's exhaustive search for a GIS for their environmental analyses found no adequate UNIX-base system in the 1980's. GRASS was built in a modular form with the cooperation and guidance of other governmental, university and commercial organizations. Thus GRASS has always been an open test bed for GIS development as well as an open general-use GIS for environmental/scientific purposes.

Participatants in this forum will examine whether and how GRASS-GIS can continue to serve as a test-bed for piece-wise development and evaluation of new approaches to GIS technology, as a prototype for interoperability, and as a vehicle for network GIS applications .

```
The following points are proposed to begin discussion:
1. How can GRASS serve users as a GIS for general use, and for testing
        technology development? What capabilities are desired?
2  a. Who has resources to implement such capabilities (in a manner
consistent
                with long term support in future realeases where
appropriate)?
   b. Can GRASS be enhanced, upgraded, bug-fixed, beta-tested, and released
        completely with in-kind resources, or is additional funding required?
   c. If additional funding is needed, how can such funding be found and
        managed?
3. Compare a Linux-like model (collaborative, decentralized) for software
        development, distribution, etc. with an ideal model for GRASS. Are
        better models available?
4. How can GRASS be managed to better serve as a public test bed for GIS
        development?
5. How can GRASS be managed to better serve: academic, public service, and
        private sector applications; commercial GIS developers; consultants
and     value-added GIS providers. How can GRASS better support international
        development and environmental conservation/management?
```


TO ALL OUR COMPANY
GIS, MODELING, GIS, MODELING
FOREVER LETS US HOLD THEIR BANNERS
RASTER VECTOR, VECTOR RASTER

Workshop on Network-Accessible Data Repositories Home Page

NCGIA Conference on GIS and Environmental Modelling

Santa Fe, New Mexico - January 21- 25, 1996

Workshop Organizers
Ray Ford, Mike Sweet, Ron Righter, Joe Glassy
University of Montana
Missoula, Montana 59812, USA

Overview

As part of the NCGIA 3rd International Conference/Workshop on GIS and Environmental Modelling in Santa Fe, New Mexico January 21-25, there will be a 2 hour workshop on Sunday afternoon, January 21st, focusing on specific technical aspects of building and maintaining a repository with environmental and natural resource information. The workshop will start with several brief presentations from various repository builders, then move into a general discussion mode. [A separate time and demonstration area will also be set up for respository builders to demonstrate the specific contents of their repositories.]

Tentative List of Presentations/Issues

(1) **R.Ford** - Computer Science, M.Sweet - School of Forestry, R.Righter - Wildlife Spatial Analysis Lab, University of Montana.

Issues

- a repository vs. an index
- object-oriented methodolgy as a basis for indexing
- linking a customized system to the Web

Additional Information

- *NCGIA Conference Paper:* M.Sweet, R.Ford, R.Righter, "A Network-Accessible Repository for the Characterization of Spatial Ecosystem Components"
- *Web Site:* University of Montana - Distributed Applications and Systems Lab: "Ecosystem Information System (EIS)"

(2) R.Kramer, **R.Nickolai**, Forschungszentrum Informatik (FZI), Computer Science Research Center, Karlsruhe, Germany.

Issues

- background info: environmental data catalogue UDK
- comparison: meta-information system vs. repository
- design recovery of an existing datamodel and system
- enhancing an existing data model

Additional Information

- *NCGIA Conference Paper:* A. Koschel, R. Kramer, R. Nicolai, et al., "A federation architecture for an environmental information system incorporating GIS, the World Wide Web, and CORBA"
- *Web Site:* FZI - Database Research Group: "Environmental Information Systems (EIS)"

(3) **L.Knapp** - IBM Government Systems, P.Andrews - USDA/US Forest Service, J.Turek - IBM TJ Watson Research Labs

Issues

- constraints imposed on the storage, retrieval, and distribution of datasets based on dataset size
- role and requirements for compression algorithms: support for both lossless and lossy representation, multiple resolutions, processing on compressed form, etc.
- impact on novel applications involving remotely sensed imagery

Additional Information

- *NCGIA Conference Paper:* L.Knapp, P.Andrews, J.Turek, "Image Navigation for Wildland Fire Location Mapping"
- *Web Site:* IBM T.J.Watson Research Labs: "Retrieval of Digital Images by Means of Content Search"

  Others to be added

PRELIMINARY (UMBRELLA) ABSTRACT:

Land Use Modeling Session and Workshop After a long and tumultuous history, the field of land use modeling is being reexamined and reinvigorated. Begun as early as the 1920s, prior land use modeling was typically approached from the perspective of the monetary economist. By the mid 1970s, land use modeling was perceived as a complex and cumbersome science and the field itself was virtually abandoned by economists and policy analysts due in large part, to internal criticism by influential modelers. Recently there is renewed interest in examining land use modeling from other perspectives, most notably with attention to understanding anthropogenic drivers of land use change and mitigating their negative effect on the environment -- inclusive of economic considerations. Loss of wetlands for increased agricultural production, deforestation, and agricultural conversion to urban land uses are examples of relevant problems facing policy makers at global, regional and local levels. Land use modeling is being reinvigorated by computing technologies such as geographic information systems and parallel processing, and by new paradigms, methods, and techniques examples of which include; ecosystem management, ecological economics, landscape ecology, and spatial modeling. The importance of land use modeling is underscored by a recent initiative of the International Global Biosphere Programme (IGBP) to elsewhere in the program (land surface change), it is the objective of this land use session and related workshop to bring together leading researchers to speak about theory and practice from different perspectives and to provide a forum in which to teach and discuss key issues. Currently three case study papers are anticipated, emphasizing different scales; one global, one regional, and one local. A lead or key talk by a recognized leader and authority in the field is planned to open the session and to provide a philosophical basis and methodological overview prior to subsequent paper presentations.

Bryan C. Pijanowski
Department of Entomology
Michigan State University
East Lansing, Michigan 48824
office: (517) 432-0039
lab   : (517) 355-4561
fax   : (517) 353-4354
email: pijanows@esalsun10.ent.msu.edu

Brad Parks
Cooperative Institute for Research in Environmental Sciences (CIRES)
Campus Box 449
University of Colorado
Boulder, Colordao 80309-0449
phone: 303-497-6330
fax: 303-497-6513
email: 74063.3124@compuserve.com

Introduction

Efficient techniques for the management of environmental data are useful not only for decision makers in government or private industry. In environmental protection, the public demand for openness and data transparency is unusually high. Recent legislation reflects this trend, especially at the European level. According to a recent EU recommendation on environmental information (modeled after the American Freedom of Information Act), almost all data stored at public agencies should be made available to any citizen on demand (Council of the European Communities1990). The few exceptions refer to privacy issues, such as the protection of industrial secrets, which need to be shielded from public view by appropriate authorization mechanisms. In general, however, public records on private companies are not exempted from this ruling.

As a result of these political developments, there is a major demand for appropriate tools to manage and aggregate environmental information distributed in a heterogenous computer network. A particular need exists for convenient navigation aids that help users to take advantage of such a distributed information system, regardless of their computer literacy. Starting from some environmental query or problem formulation, such a navigation aid should help users to localize the relevant data sets and to retrieve them quickly and in a user-friendly manner. An essential prerequisite for both navigation and data transfer is a *metainformation system* that contains data about the format and the contents of the available data.

The *UDK (Environmental Data Catalogue)* is such a metainformation system and navigation tool. It documents collections of environmental data from the government and other sources. These data sets may be available either online or by request to the responsible data administrator. Potential users of the system include government agencies, industry, as well as the general public. The UDK helps them to get answers to the following questions:

- Which relevant information is principally available for a given problem?
- Where is this information stored?
- How can this information be retrieved, and how long does it take?

The UDK design presented in this paper is the result of several years of research and development (Lessing and Schmalz 1994, Lessing and Schütz1994, Lessing1989). In 1990, the Environmental Ministry of the State of Lower Saxony launched a research project with funding from the German Federal Environmental Protection Agency. Two years later, an international working group was formed to oversee the UDK design and its further development into a practical software tool. In 1994, Austria passed an Environmental Information Law that introduced the UDK as the official navigation tool for all environmental information on record. In January 1995, a first version of the UDK was made available in Austria; most German states will follow later this year. The UDK is currently also under evaluation by several other European countries, including Switzerland, Italy, Sweden, and Norway.

Next Up Previous

Next: The UDK Object **Up:** UDK: A European Environmental **Previous:** UDK: A European Environmental

---

Oliver Günther

Next Up Previous

Next: UDK Object Classes **Up:** UDK: A European Environmental **Previous:** Introduction

The UDK Object Model

The UDK is based on a *three-way object model* that distinguishes between environmental objects, environmental data objects, and UDK (meta) objects (Fig. 1). The term *environmental object* is used to describe the real-world objects making up the environment. This includes natural entities, such as lakes and biotopes, as well as man-made objects, such as factories or highways. Each environmental object is described by a collection of *environmental data objects*. There are a variety of ways to obtain such a description, including measurements, observations, but also value judgements. A typical environmental data object is a series of measurements that captures the concentration of a certain substance in a river (the corresponding environmental object).

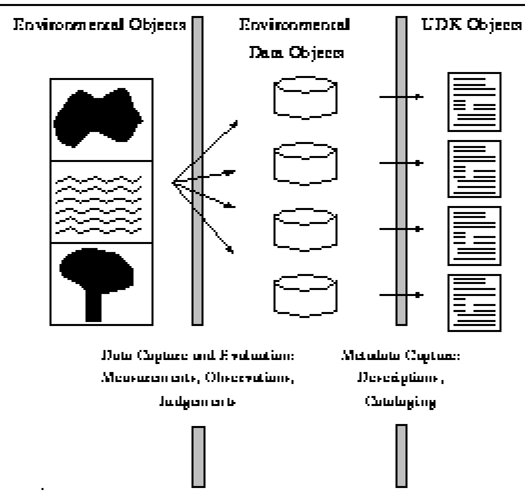


Figure 1: Three-way object model of the UDK

According to a recent definition (Swoboda et al. 1995), there are two criteria an environmental data object should fulfill:

1. *Environmental Context:* Environmental data objects are information that refers to real, material components of the environment (environmental objects), or to processes or projects that are related to such environmental objects;
2. *Data Capture:* Environmental data objects are information that is available in the form of a written text, sound, or images (analog or digital). It has been obtained by one of the following ways of data capture:
    - descriptive;
    - empirical;

- by collection and aggregation of other data;
- by a simulation or some other modeling process;
- by value judgements.

Each environmental data object is in turn associated with exactly one metadata object that specifies its format and contents. On the screen, each such *UDK object* is represented by one or more masks; see Figure 2 for an example.

Figure 2: A mask representing a UDK object

UDK objects may exist for environmental data objects at various aggregation levels simultaneously. Consider, for example, a national groundwater database that contains a large number of measurements from all over the country. There is one UDK object representing this database as a whole. In addition, however, there may be one UDK object each for the measurements from a certain county, there may be UDK objects representing the measurements from a particular station, and there may even be UDK objects that represent single measurements. There may also be UDK objects for groupings that are orthogonal to this aggregation hierarchy, such as UDK objects representing the measurements that were taken in a given month. The decision to create an object is based on a cost/benefit analysis, depending on the particular applications a user has in mind. An important issue in this context is the maintenance of the semantic associations between those UDK objects and the corresponding environmental data objects; see Section 4 for further details.

Up to now, UDK objects have been identified by their position in the *primary tree*, a directed graph whose nodes correspond to the UDK objects and whose edges represent responsibilities of agencies and departments for particular sets of UDK objects, as well as part-of-relationships between large data collections (e.g. a groundwater database) and their components (e.g. the data sets corresponding to particular measuring stations). This approach to identify objects is unsatisfactory for a variety of reasons. Most importantly, UDK objects may lose their identity when they are relocated in the primary tree due to some reorganization (such as the transfer of a department from one ministry to another). In this case, the objects that were relocated have to be recreated under a new ID at the new location. As an alternative, we are currently investigating the possibility to use *object identifiers (OIDs)*, a concept well-known from the domain of object-oriented databases. OIDs are created by the system; they are usually not visible by the user. To guarantee universal uniqueness, the generation of the OID is usually based on the CPU number, as well as the current date and time-of-day.

Next Up Previous

Next: UDK Object Classes **Up:** UDK: A European Environmental **Previous:** Introduction

---

Oliver Günther

Next Up Previous

Next: Semantic Associations Between **Up:** UDK: A European Environmental **Previous:** The UDK Object

UDK Object Classes and Inheritance

In order to structure the wide variety of UDK objects, and to facilitate both their capture and their administration, we recently presented a first proposal for a *UDK class concept* (Lessing, Günther, and Swoboda 1995). Here we distinguish between seven classes of environmental data objects:

1. project data (construction projects, environmental impact studies, etc.)
2. empirical data (measuring series, laboratory data, etc.)
3. data about facilities (factories, buildings, etc.)
4. maps
5. expertises and reports
6. product data
7. model data (simulations etc.)

For each of these seven classes there is a UDK class that contains the describing UDK objects. Each UDK class corresponds to a mask that is used for the capture and administration of the corresponding UDK objects. The basis for this pragmatic proposal were the user requirements that came up during the first months of UDK data capture. Obviously, this classification needs to be reviewed and possibly extended from time to time in order to reflect the change of user requirements. We feel it is important, however, that this top-level classification reflects a consensus of all UDK participants.

Another extension that is currently planned concerns the *vertical* structure of this classification. In particular, we intend to turn this flat class structure into an object-oriented class hierarchy that allows the inheritance of object attributes. The hierarchy should be structured as follows.

- The root of the hierarchy (level 0) consists of the generic class *UDK_Object* with four obligatory attributes: the unique object identifier (OID), the object name, the date when the object was last modified, and the agency (or the person) that is responsible for the object. Optional attributes, such as a textual description, may be included as well. Note that this generic class is not an abstract class, i.e., it may contain objects that are not included in any of its subclasses.
- Level 1 contains a relatively small number of classes that represent a consensus between all UDK participants. Currently, this level would correspond to the seven classes described above. Changes at this level are subject to negotiation between the UDK member countries.

- On the subsequent levels of the hierarchy, participating countries or agencies are free to introduce additional subclasses depending on their particular requirements. This kind of flexibility is important not only for efficiency reasons. It is also crucial in order to secure acceptance for the UDK throughout its intended user community, especially in government agencies at the national and the local level.

Class attributes are inherited along this class hierarchy in an object-oriented manner. This includes the possibility to upgrade selected attributes from being optional to being required. It also means that attributes that are specific to a certain subclass, but not to its superclass(es), can be masked out when looking only at the superclass. For example, consider a particular topographic map m and its UDK object U_m. m is an element of the class *topographic_map*, which is a subclass of the class *map*. If one now looks at the UDK object U_m through the mask corresponding to the class *map*, one only sees the attributes of *map*. The additional attributes that may have been introduced to describe *topographic* maps, as opposed to general maps, are not visible in this case.

This feature, which is typical for object-oriented environments, is a crucial element of standardization in the presence of application-specific extensions on the class hierarchy levels 2 and below. Any tool that is supposed to work at the national (or international) level across particular agencies or user communities can rely on the availability of the attributes defined at level 1. Maintenance and version management are other issues that need to rely on a stable class and attribute structure at the higher levels of the object hierarchy. It is therefore important to take organizational and technical precautions to make sure that users observe this principle in the presence of user-specific extensions and the resulting complexity in the class structure. The technical details of the implementation of these lower hierarchy levels are still subject to discussion.

Next Up Previous

Next: Semantic Associations Between **Up:** UDK: A European Environmental **Previous:** The UDK Object

---

Oliver Günther

Next Up Previous

Next: Conclusions and Future **Up:** UDK: A European Environmental **Previous:** UDK Object Classes

Semantic Associations Between UDK Objects

Orthogonal to the class hierarchy described in the previous section, the UDK offers users the possibility to connect *concrete UDK objects* with each other in a hypertext fashion. The resulting structures are directed graphs whose nodes correspond to UDK objects and whose edges represent semantic associations between them or between their respective environmental data objects. The semantics of those edges may vary; we will later propose a type system for edges to make this aspect more explicit. Note that those semantic nets are completely independent of the class hierarchy described in the previous section. While the nodes of the class hierarchy are *UDK object classes*, the nodes of the structures described in the sequel represent *concrete UDK objects*.

The most important such graph structure is the *primary tree* or *primary catalogue*. Each UDK object corresponds to exactly one node of this tree structure, i.e., there is a 1:1 relationship between primary tree nodes and UDK objects. The links in the upper part of the tree serve to represent responsibilities of agencies and departments for particular sets of UDK objects. The agency that is in charge of a UDK object has to make sure that its information is correct and up-to-date. It is also responsible for the creation and deletion of UDK objects in the associated subtree(s). In the lower part of the tree, the links are used to represent part-of-relationships between large data collections (e.g. a groundwater database) and their components (e.g. the data sets corresponding to particular measuring stations). An example is given in Figure 3 that depicts the UDK objects related to a groundwater database. Here the solid arrows make up the primary tree; their semantics varies between ``is-responsible-for'' (in the upper part of the tree) and ``is-an-aggregation-of'' (in the lower part).

Depending on particular user requirements, there may also be *secondary catalogues* to represent other semantic associations. Like the primary tree, a secondary catalogue is a directed graph whose nodes each correspond to exactly one UDK object. Other than in the case of the primary tree, however, the resulting structure does not have to be a tree anymore. Note also that a UDK object can be referenced by any number of secondary catalogues.

A typical application of secondary catalogues concerns the representation of additional aggregation relationships that are not represented in the primary tree. In Figure 3 these kind of associations are pictured as dotted arrows. These kind of links are often useful in order to refer users to relevant aggregated data sets first before, upon request, giving them access to more detailed data. Another application of secondary catalogues is the construction of personal association structures. The ``debate'' association in Figure 3 (dashed line) is an example for such a structure. For these free structures the system does not require users to restrict themselves to a tree structure. Similar to the freedom one has for linking pages in the World Wide Web, any directed graph structure is permitted, including graphs with cycles. The idea is to give UDK users a maximum amount of flexibility to connect and associate the various information items making up their working environment. With an attractive user interface, this option should be of great interest to a large group of users. What is important is that it has to be reasonably easy to create personal UDK objects and links. Furthermore, it is essential that those ``personal'' structures can be isolated from the public part of the UDK, so users can build confidential structures that are visible just for them or for their team.

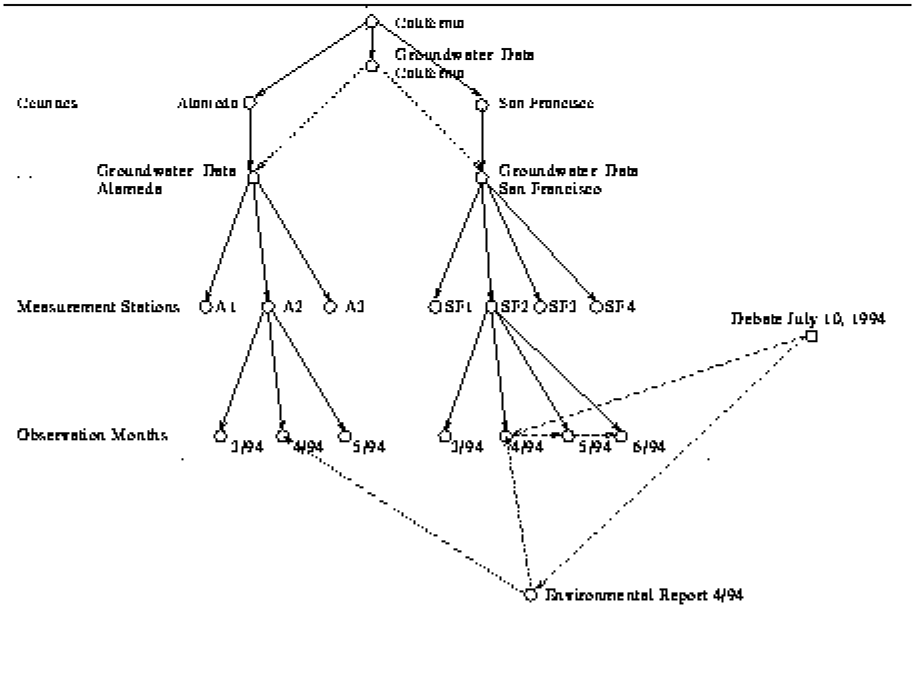


Figure 3: A selection of UDK objects and their associations in the context of a groundwater database

In summary, it is important to note that the links connecting UDK objects may have a great variety of semantics. These different types of links need to be made explicit in the UDK by a labeling scheme. Users should have the option to choose the types of links they want to see at a given time. This would allow them to see a UDK object in a variety of contexts and to switch back and forth between those different representations. On the screen this could be supported, for example, by different colors and drawing modes for different types of links (Fig. 3).

Next Up Previous

Next: Conclusions and Future **Up:** UDK: A European Environmental **Previous:** UDK Object Classes

Oliver Günther

Next Up Previous

Next: References **Up:** UDK: A European Environmental **Previous:** Semantic Associations Between

Conclusions and Future Work

The UDK is a metainformation system and navigation tool that documents collections of environmental data from the government and other sources. Given the extreme success of the World Wide Web (WWW), we expect a significant amount of this kind of data to be available via WWW in the very near future. At this point there is no question that the Web is the most promising option to follow the spirit of the EU guideline and to make environmental information *really available* to anybody who is interested.

The UDK could play a major role in helping users to navigate in this overwhelming information pool, to identify which data is relevant for a given query, and to retrieve it fast and in a user-friendly manner.Both Austria and the German state of Baden-Württemberg have recently commissioned WWW implementations of the UDK, and we expect prototypical versions to be available on the Web by early 1996. The technical realization is straightforward. Each UDK object can be turned into a Web page. HTML links can then be used to implement primary and secondary catalogues and to establish connections to the actual environmental data objects. (A slightly different approach was presented by Kramer and Quellenberg (1996) and Koschel et al. (1996)). Given the openness of the Web paradigm, the integration of most relevant data formats (such as postscript documents, sounds, or images) is easily possible.

Oliver Günther

References

Council of the European Communities (1990, July). Council Directive (90/313/EEC) of 7 June 1990 on the freedom of access to information on the environment. *Official Journal of the European Communities L158*, 56-58.

Koschel, A., R. Kramer, R. Nikolai, W. Hagg, and J. Wiesel (1996). A federation architecture for an environmental information system incorporating GIS, the World Wide Web, and CORBA. In this volume.

Kramer, R. and T.Quellenberg (1996). Global access to environmental information. In R. Denzer, D. Russel, and G. Schimak (Eds.), *Proc. 1995 International Symposium on Environmental Software Systems*, International Federation for Information Processing (IFIP), London, pp.209-218. Chapman and Hall.

Lessing, H. (1989). Umweltinformationssysteme - Anforderungen und Möglichkeiten am Beispiel Niedersachsens. In A. Jaeschke, W. Geiger, and B. Page (Eds.), *Informatik im Umweltschutz*, Berlin/Heidelberg/New York. Springer-Verlag.

Lessing, H., O. Günther, and W. Swoboda (1995). Ein objektorientiertes Klassenkonzept für den Umwelt-Datenkatalog (UDK). In *Space and Time in Environmental Information Systems*, Marburg. Metropolis-Verlag.

Lessing, H. and R. Schmalz (1994). Der Umwelt-Datenkatalog Niedersachsens. In A. Engel (Ed.), *Umweltinformationssysteme in der öffentlichen Verwaltung* . R. v. Decker's Verlag.

Lessing, H. and T. Schütz (1994). Der Umwelt-Datenkatalog als Instrument zur Steuerung von Informationsflüssen. In L. Hilty, A. Jaeschke, B. Page, and A. Schwabl (Eds.), *Informatik für den Umweltschutz*, Marburg. Metropolis-Verlag.

Swoboda, W., H. Lessing, P. Grolimund, and O. Günther (1995). Metadatenklassen im Umwelt-Datenkatalog (UDK). In F. Huber-Wäschle, H. Schauer, and P. Widmayer (Eds.), *Proc. GISI 95*, Berlin/Heidelberg/New York. Springer-Verlag.

Oliver Günther

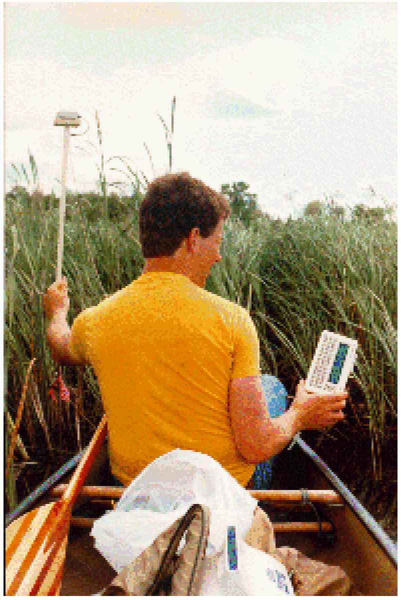


Figure 1. Taking GPS readings of a field sample point by canoe in the Pokegama wetland.

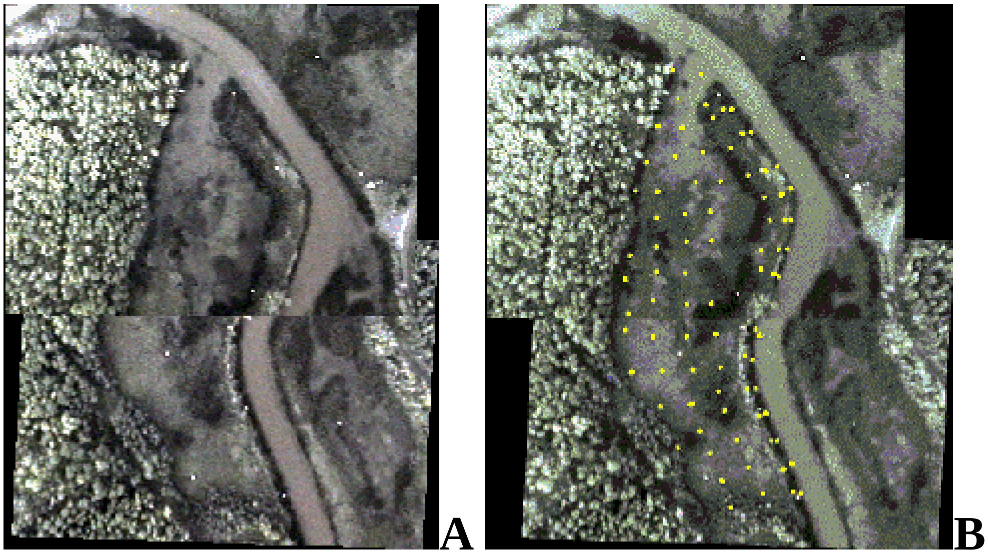


Figure 2. (A) Mosaicked, georeferenced video images taken September 1994, without attitude correction. White squares visible within the image are styrofoam targets placed in the field. (B) Mosaicked video images with attitude correction provided by the Ashtech 3DF ADU global positioning system. Yellow squares are the locations of field sample points, determined by field GPS measurements, superimposed on the video image.

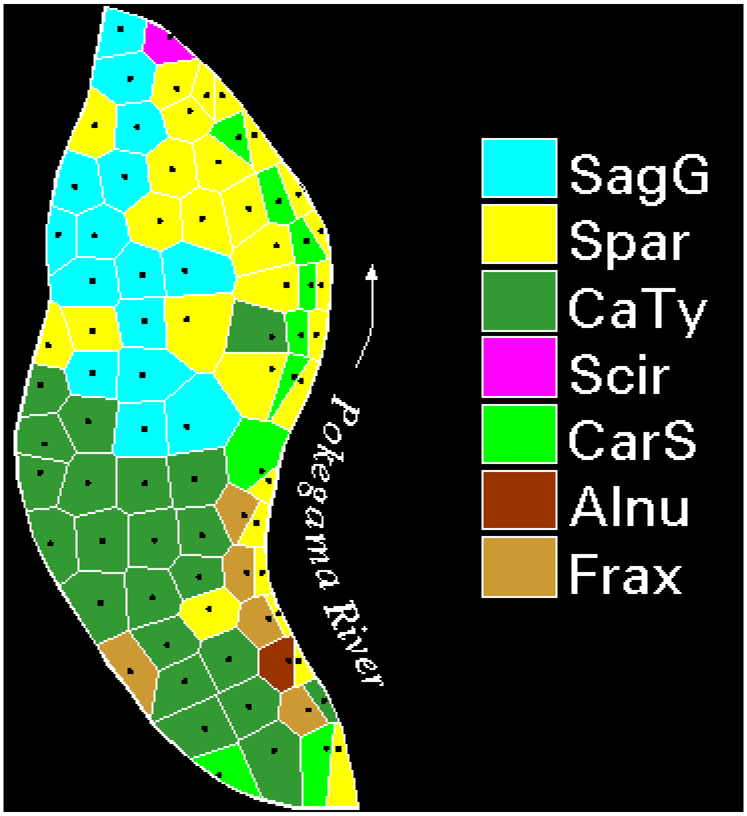
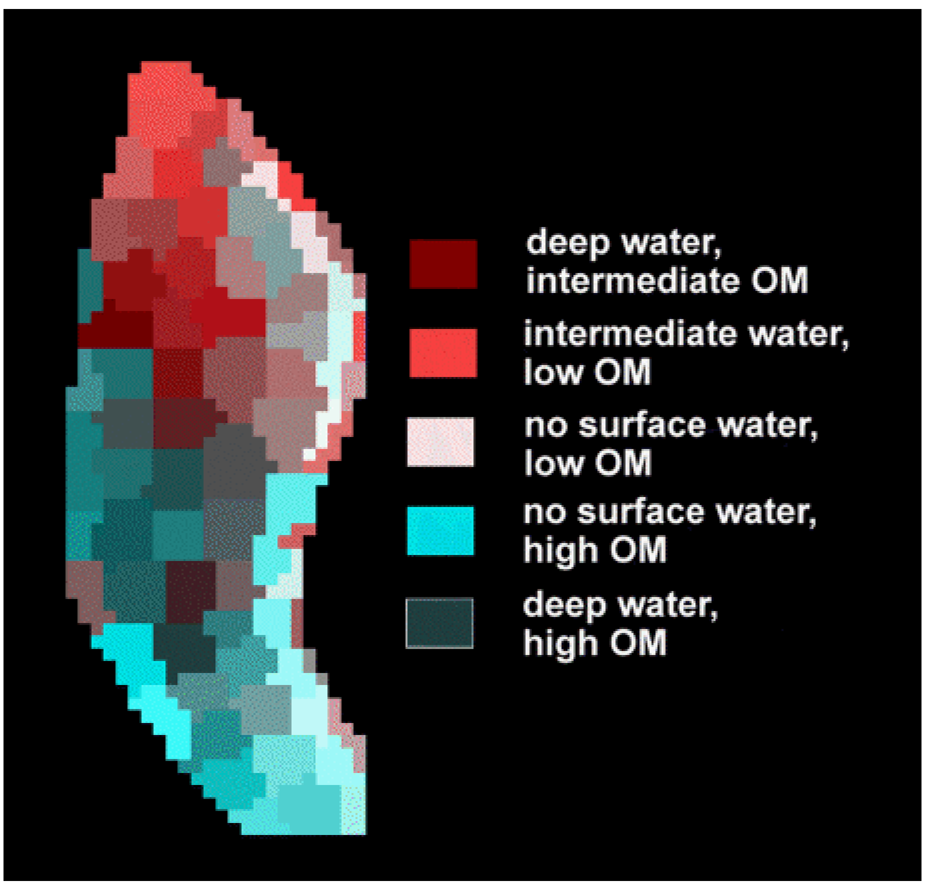


Figure 3. Thiessen polygon map based on TWINSPAN analysis of vegetation data. Alnu = *Alnus crispa*, CarS = *Carex stricta*, CaTy = *Carex/Typha*, Frax = *Fraxinus nigra*, SagG = *Sagittaria graminea*, Scir = *Scirpus americanus*, Spar = *Sparganium eurycarpum*.

A


deep water,
intermediate OM
intermediate water,
low OM
no surface water,
low OM
no surface water,
high OM
deep water,
high OM

B

Figure 4. Color composite maps of CCA results. (A) Composite map for CCA axes 1,2, and 3. (B) Composite map for CCA axes 1 and 2 only. Values for CCA axis 2 are indicated by red hues, and values for CCA axis 1 are indicated by cyan hues.

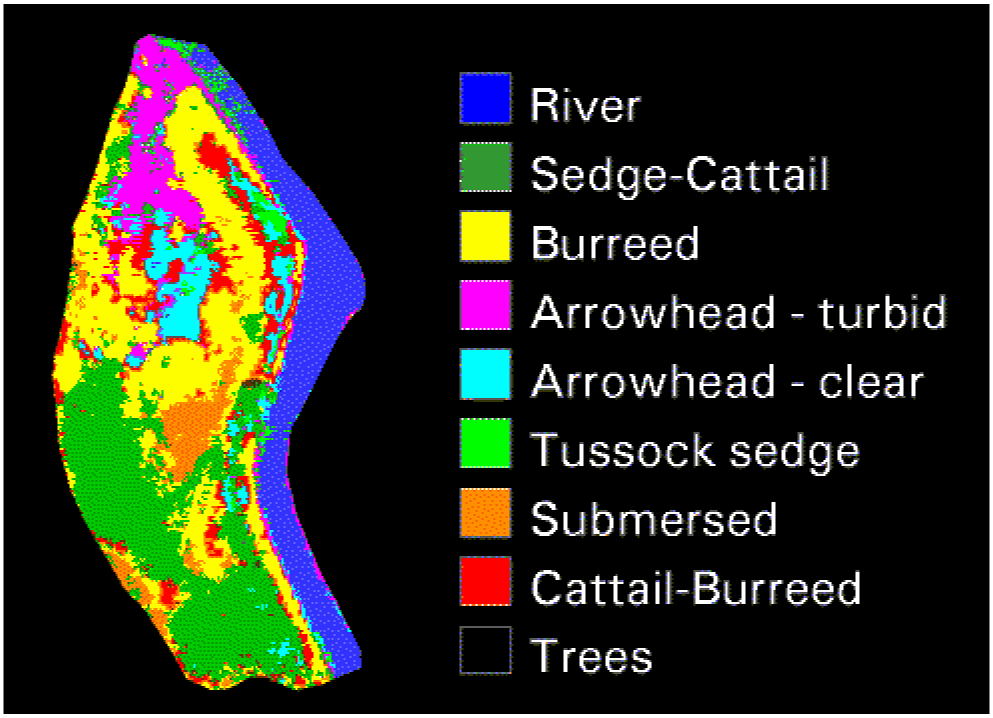
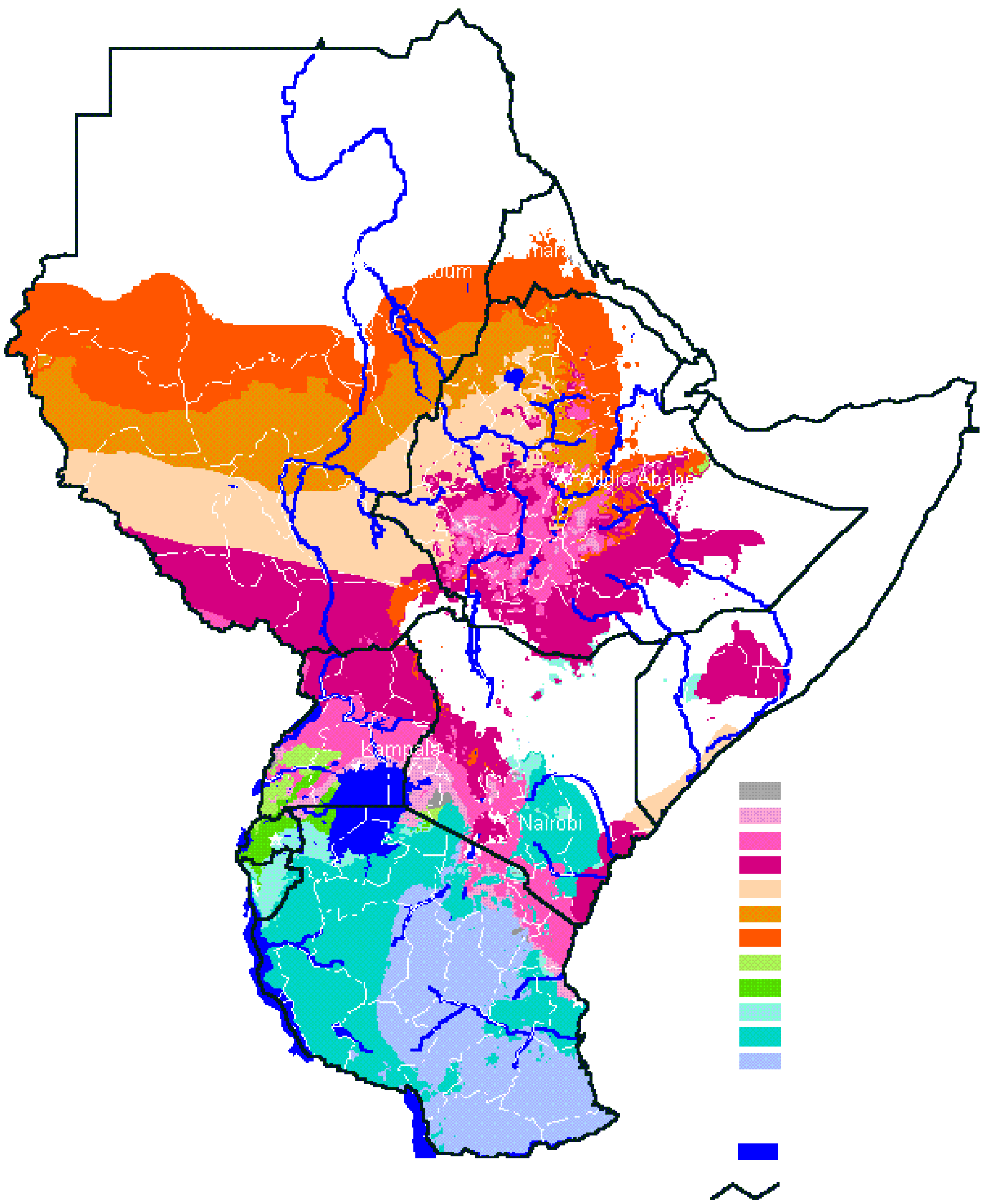
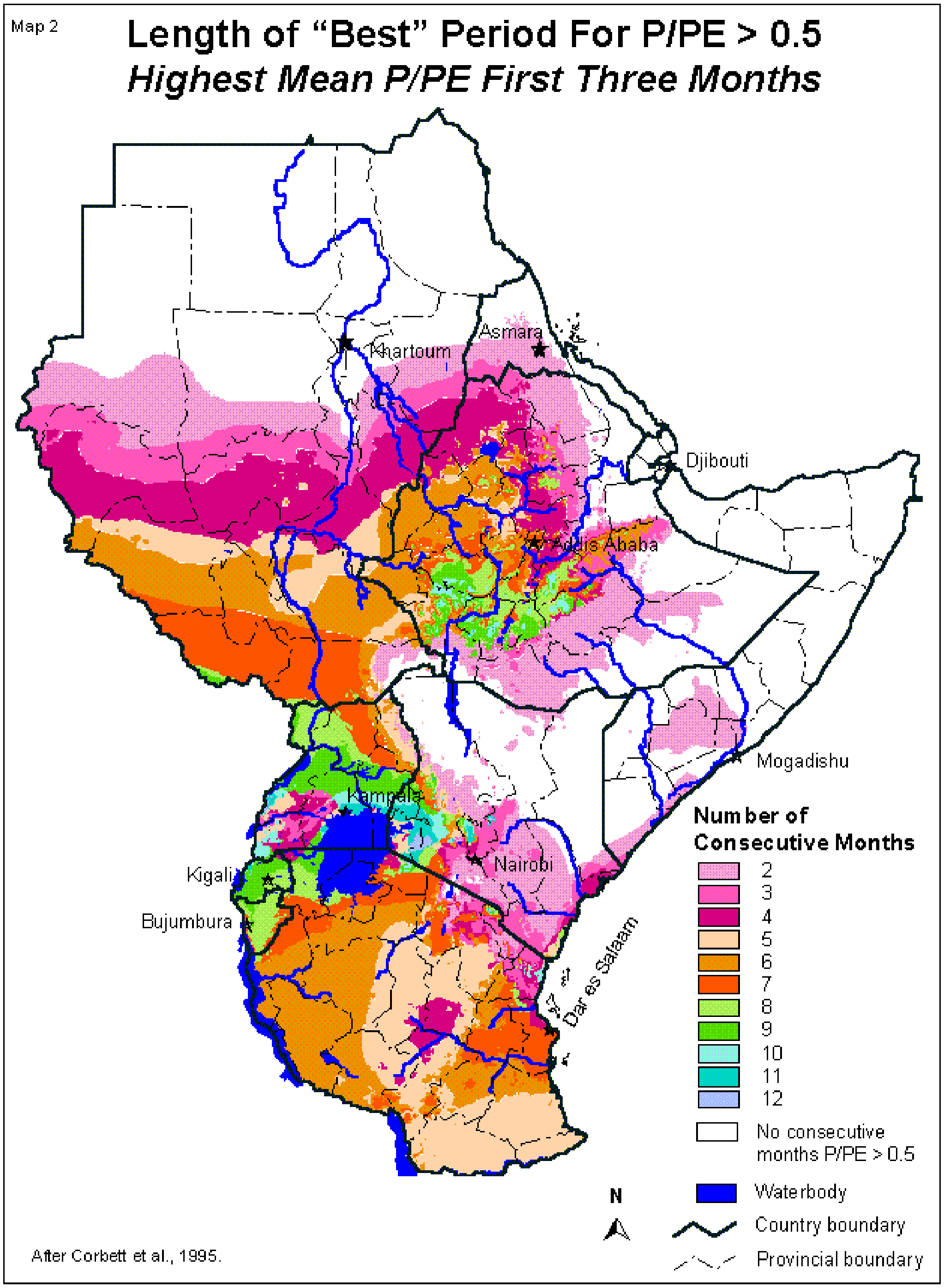
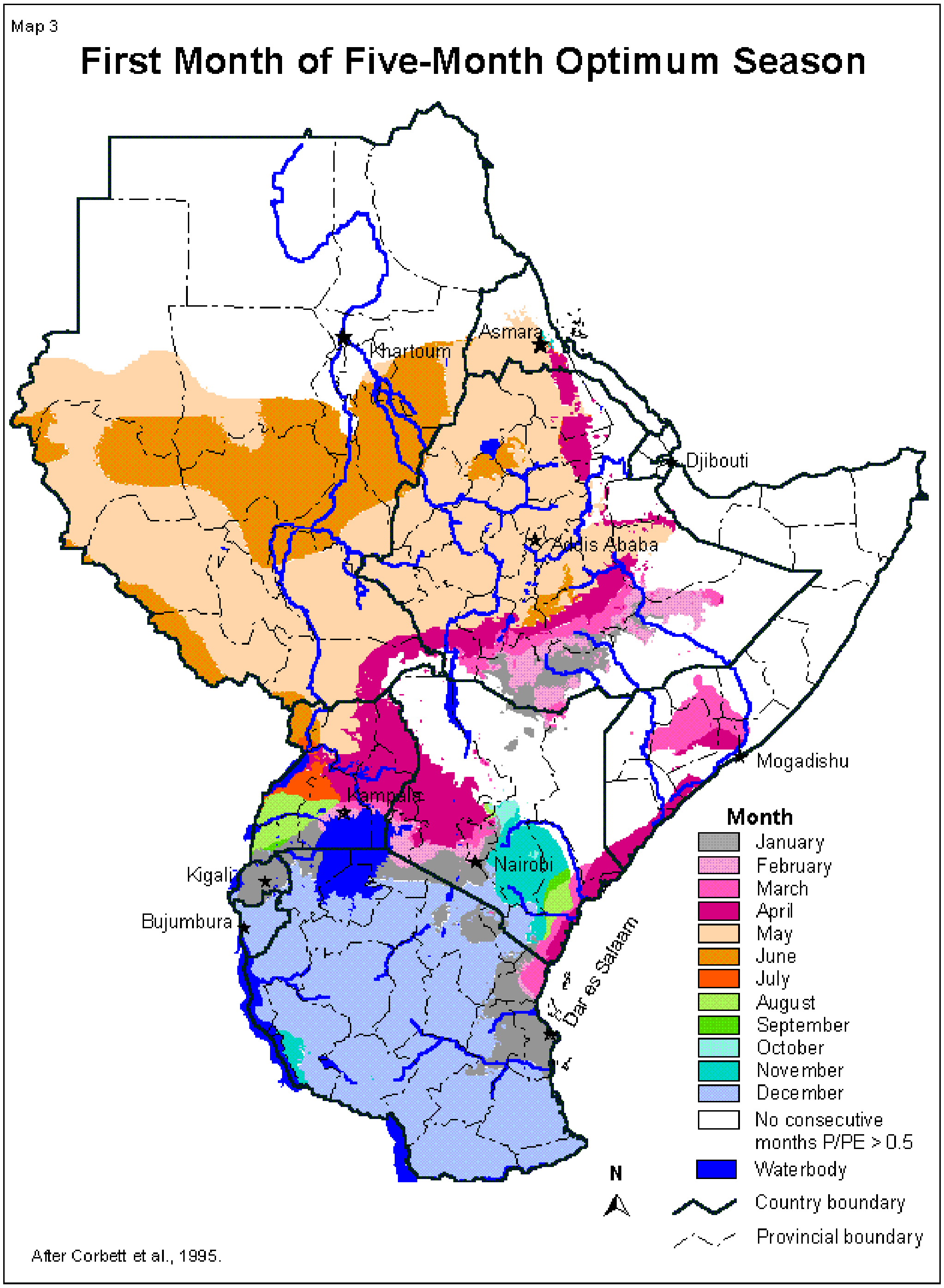


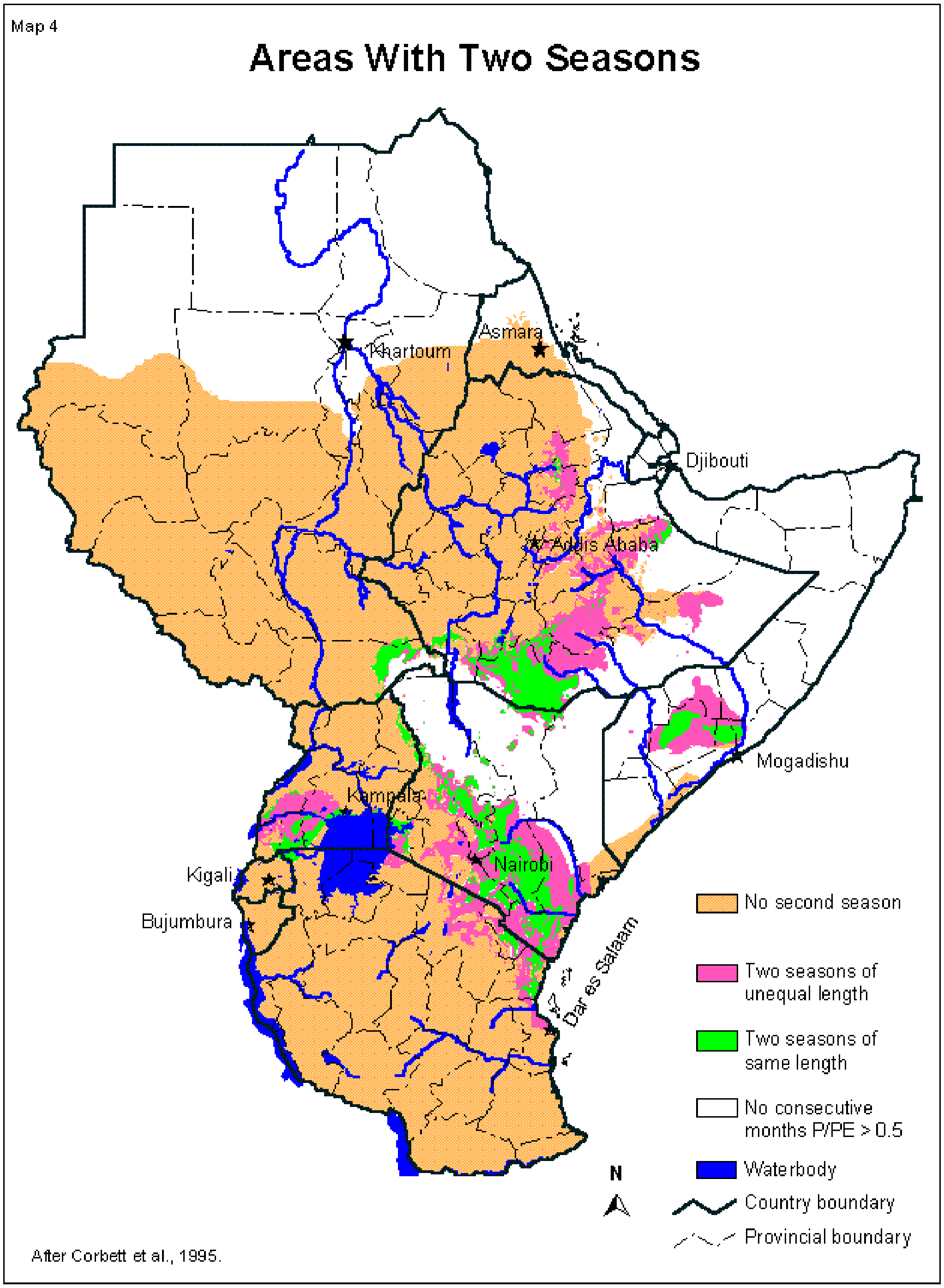
Figure 5. Vegetation map derived from multi-temporal airborne videography.

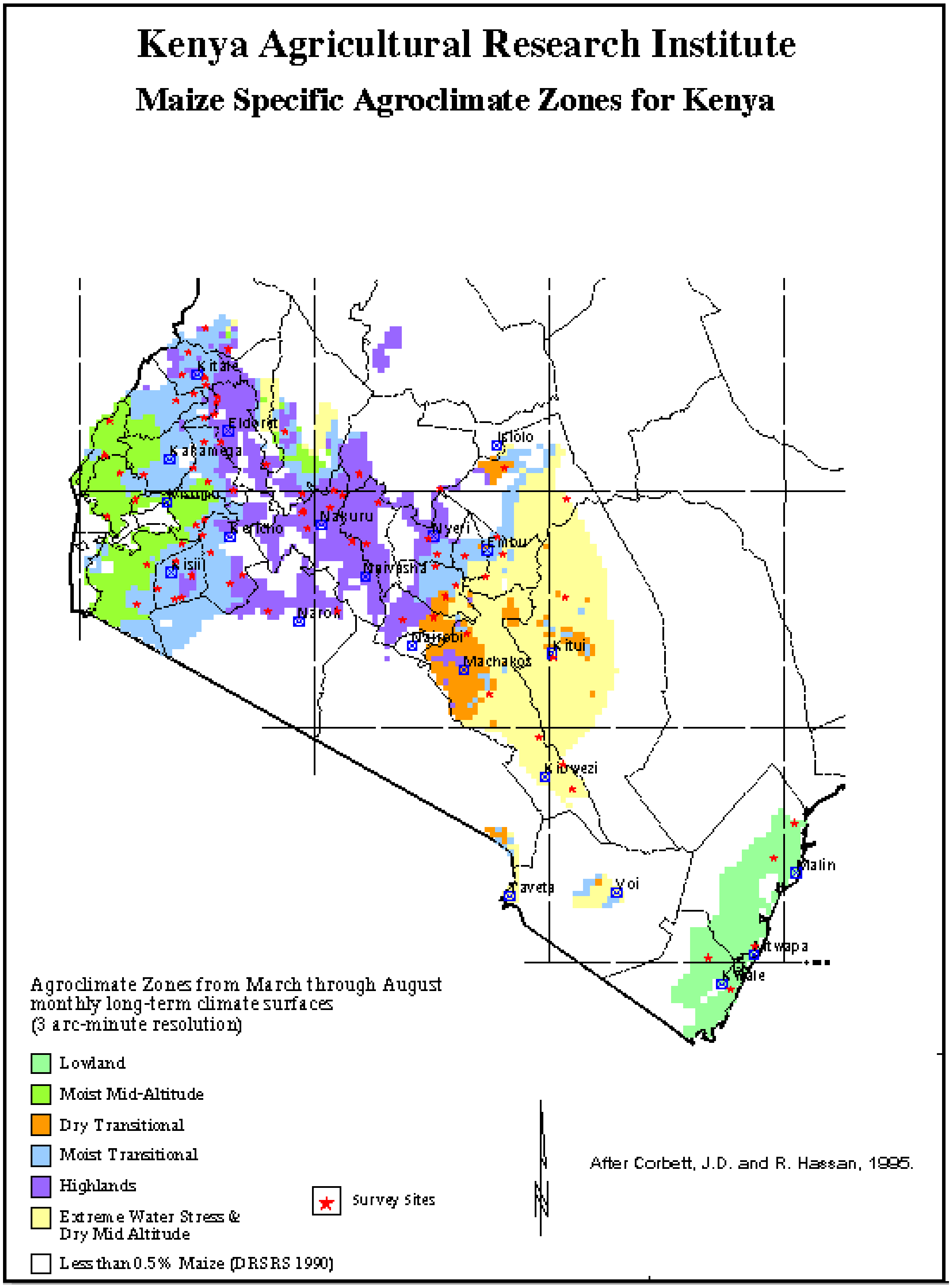

Kampala
Nairobi


Map 2
Length of "Best" Period For P/PE > 0.5
Highest Mean P/PE First Three Months
Khartoum
Asmara
Djibouti
Addis Ababa
Mogadishu
Kampala
Nairobi
Kigali
Bujumbura
Dar es Salaam
Number of Consecutive Months
2
3
4
5
6
7
8
9
10
11
12
No consecutive months P/PE > 0.5
Waterbody
Country boundary
Provincial boundary
N
After Corbett et al., 1995.

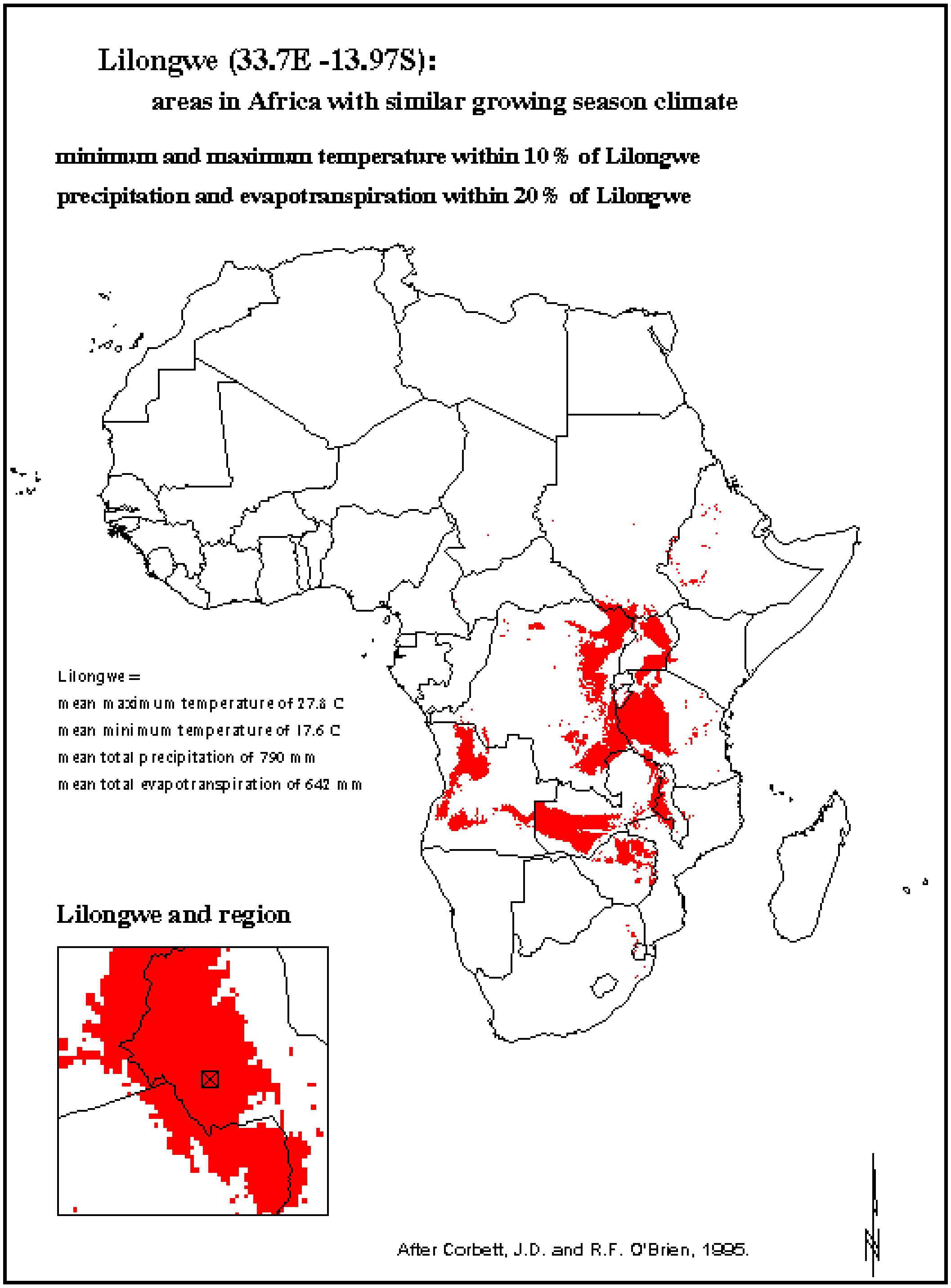


Map 3
First Month of Five-Month Optimum Season
Asmara
Khartoum
Djibouti
Addis Ababa
Mogadishu
Kampala
Nairobi
Kigali
Bujumbura
Dar es Salaam
Month
January
February
March
April
May
June
July
August
September
October
November
December
No consecutive months P/PE > 0.5
Waterbody
Country boundary
Provincial boundary
N
After Corbett et al., 1995.


Map 4
Areas With Two Seasons
Asmara
Khartoum
Djibouti
Addis Ababa
Mogadishu
Kampala
Nairobi
Kigali
Bujumbura
Dar es Salaam
No second season
Two seasons of unequal length
Two seasons of same length
No consecutive months P/PE > 0.5
Waterbody
Country boundary
Provincial boundary
N
After Corbett et al., 1995.



Kenya Agricultural Research Institute
Maize Specific Agroclimate Zones for Kenya
Kitale
Eldoret
Kakamega
Kisumu
Kericho
Nakuru
Nyeri
Embu
Kisii
Naivasha
Narok
Nairobi
Machakos
Kitui
Isiolo
Kibwezi
Taveta
Voi
Agroclimate Zones from March through August monthly long-term climate surfaces (3 arc-minute resolution)
Lowland
Moist Mid-Altitude
Dry Transitional
Moist Transitional
Highlands
Extreme Water Stress & Dry Mid Altitude
Less than 0.5% Maize (DRSRS 1990)
Survey Sites
N
After Corbett, J.D. and R. Hassan, 1995.


Lilongwe (33.7E -13.97S):
areas in Africa with similar growing season climate
minimum and maximum temperature within 10 % of Lilongwe
precipitation and evapotranspiration within 20 % of Lilongwe
Lilongwe =
mean maximum temperature of 27.8 C
mean minimum temperature of 17.6 C
mean total precipitation of 790 mm
mean total evapotranspiration of 642 mm
Lilongwe and region
After Corbett, J.D. and R.F. O'Brien, 1995.
N

Using Barts data to make road maps

This is a prototype interface, which demonstrates how to manipulate Bartholow data with Arc/Info to generate road maps for your special needs.

Select a (major) road feature

Feature: Motorway In colour: Blue

Select A road features (multi-selectable)

Features:
A Road Primary Trunk Dual C/W
A Road Primary Trunk Single C/W
A Road Primary Trunk Passing Places

In colour: Green

Select B road or other features (multi-selectable)

Features:
B Road Dual Carriage Way
B Road Single Carriage Way
B Road with Passing Places

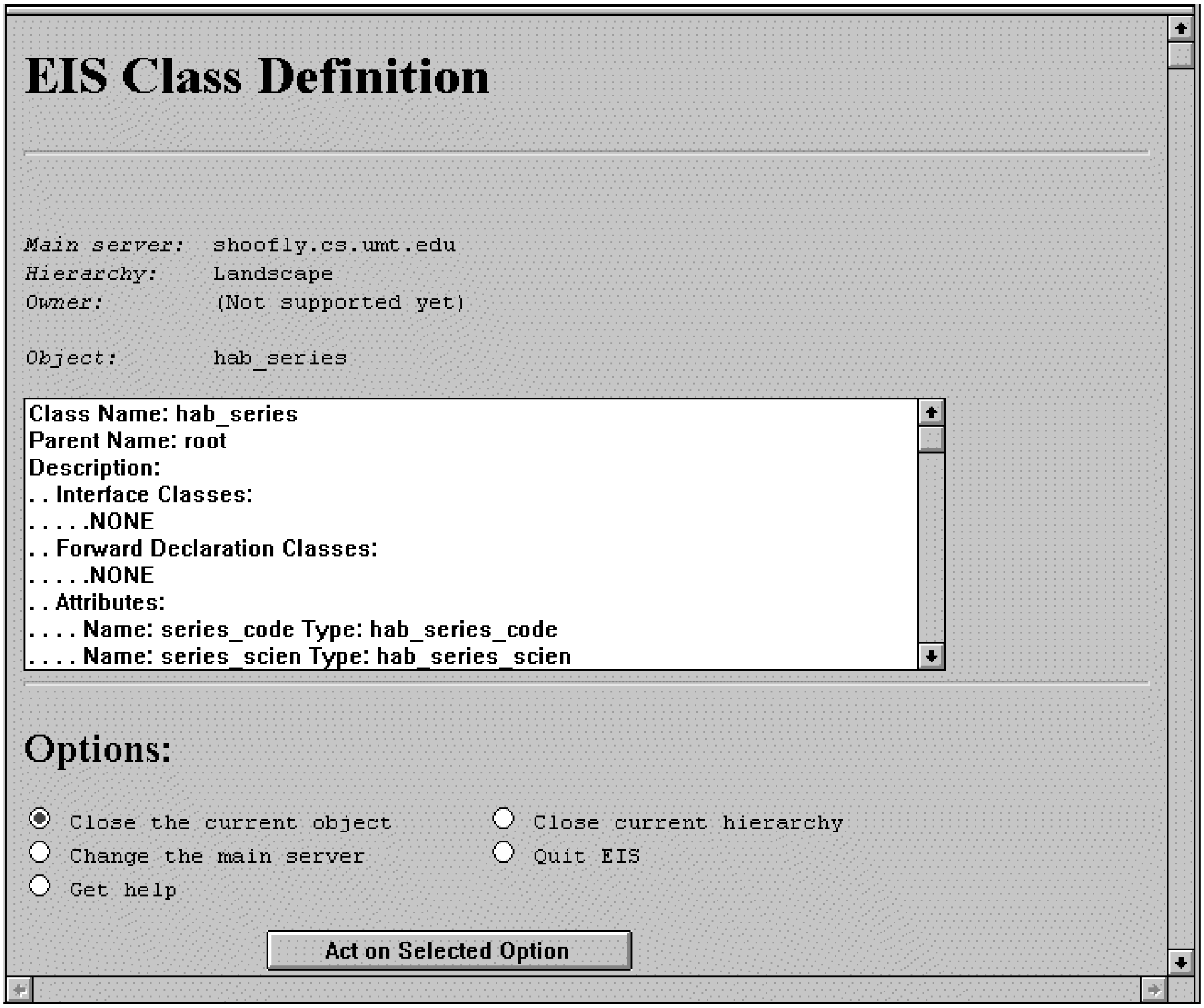
In colour: Red

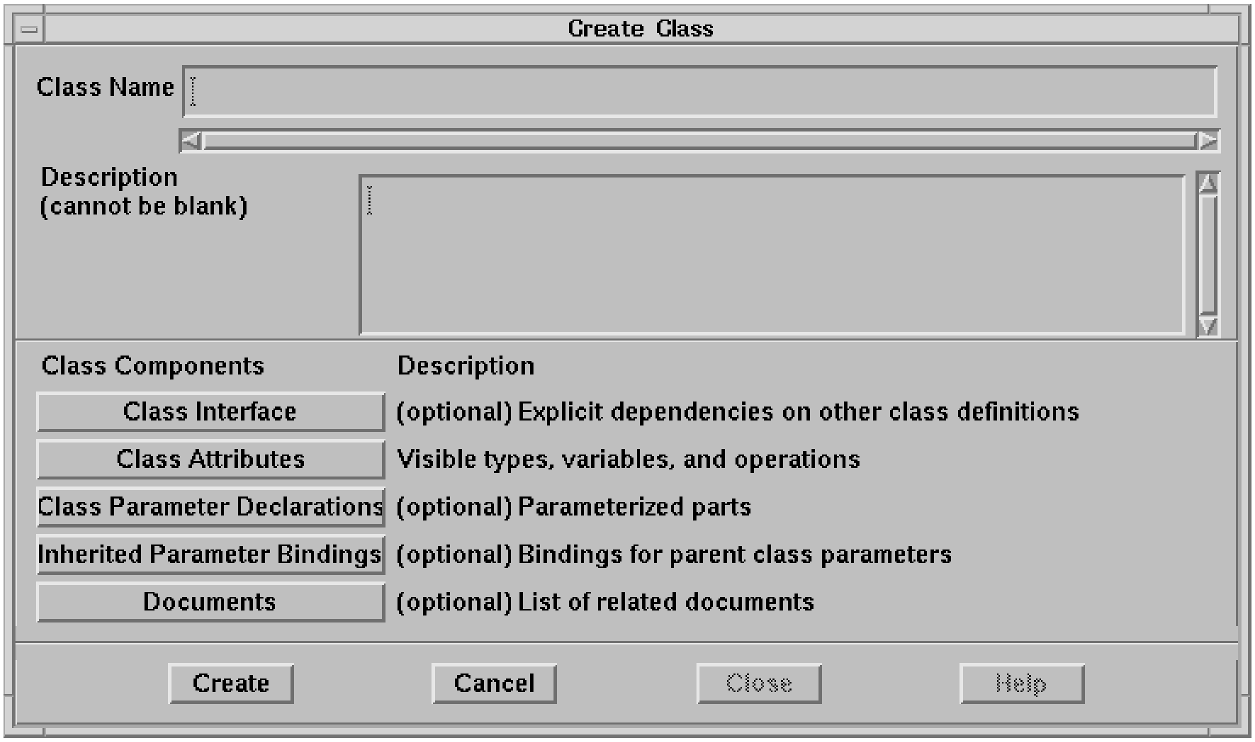
Select an additional theme (if exists) as the background

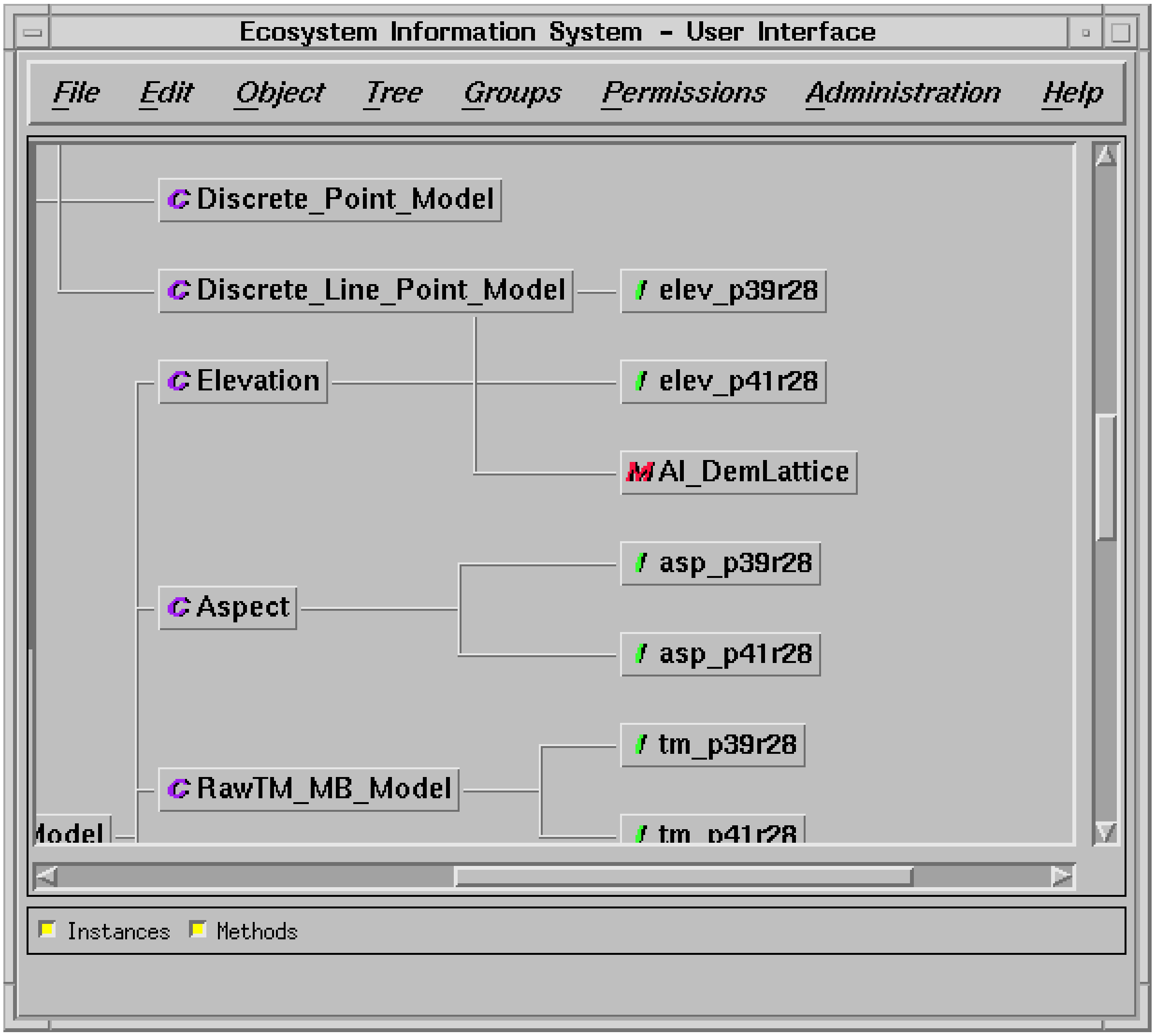
Theme: Admin Boundaries In colour: Cyan

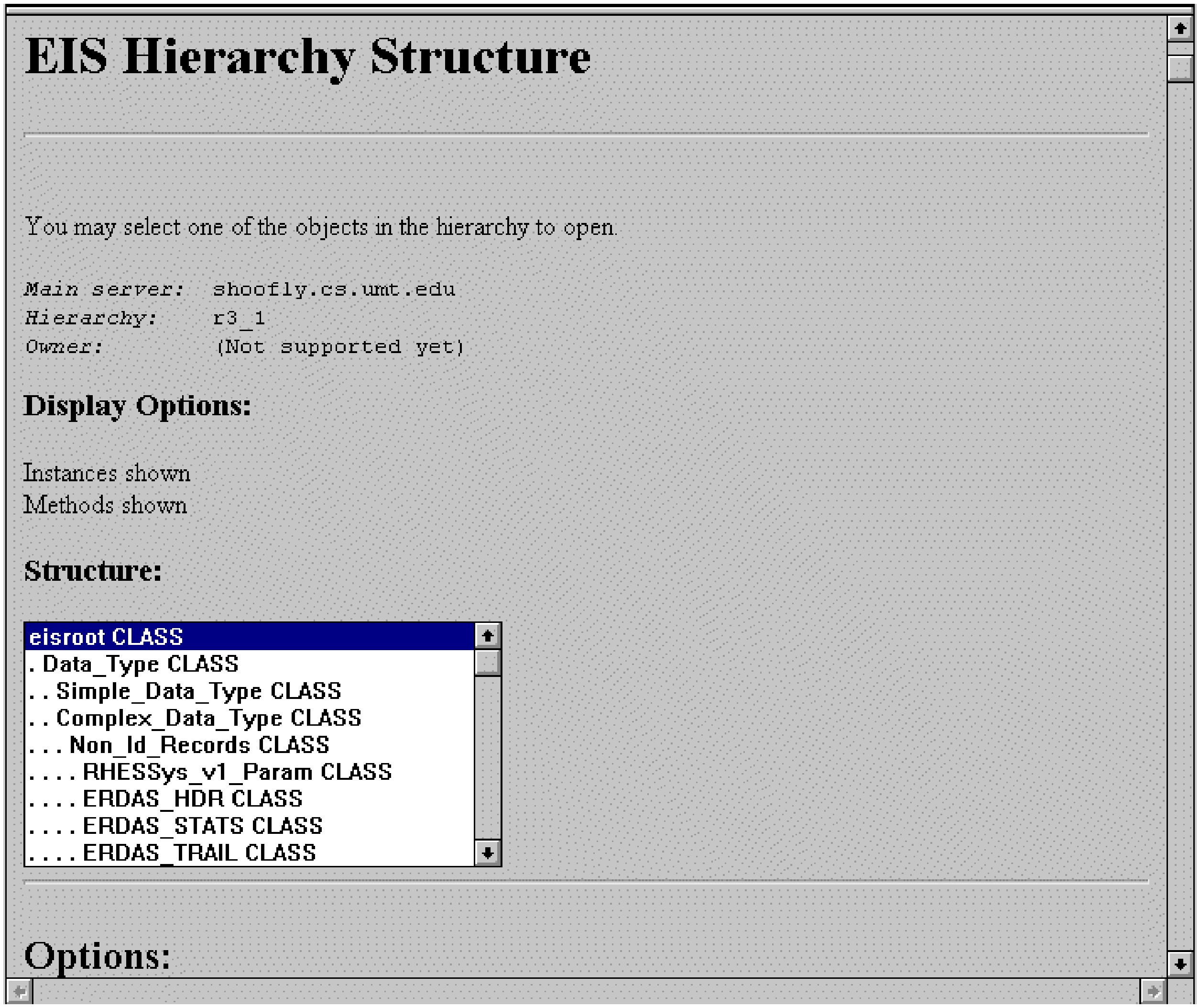
To start processing, press the button: Make a map **.**

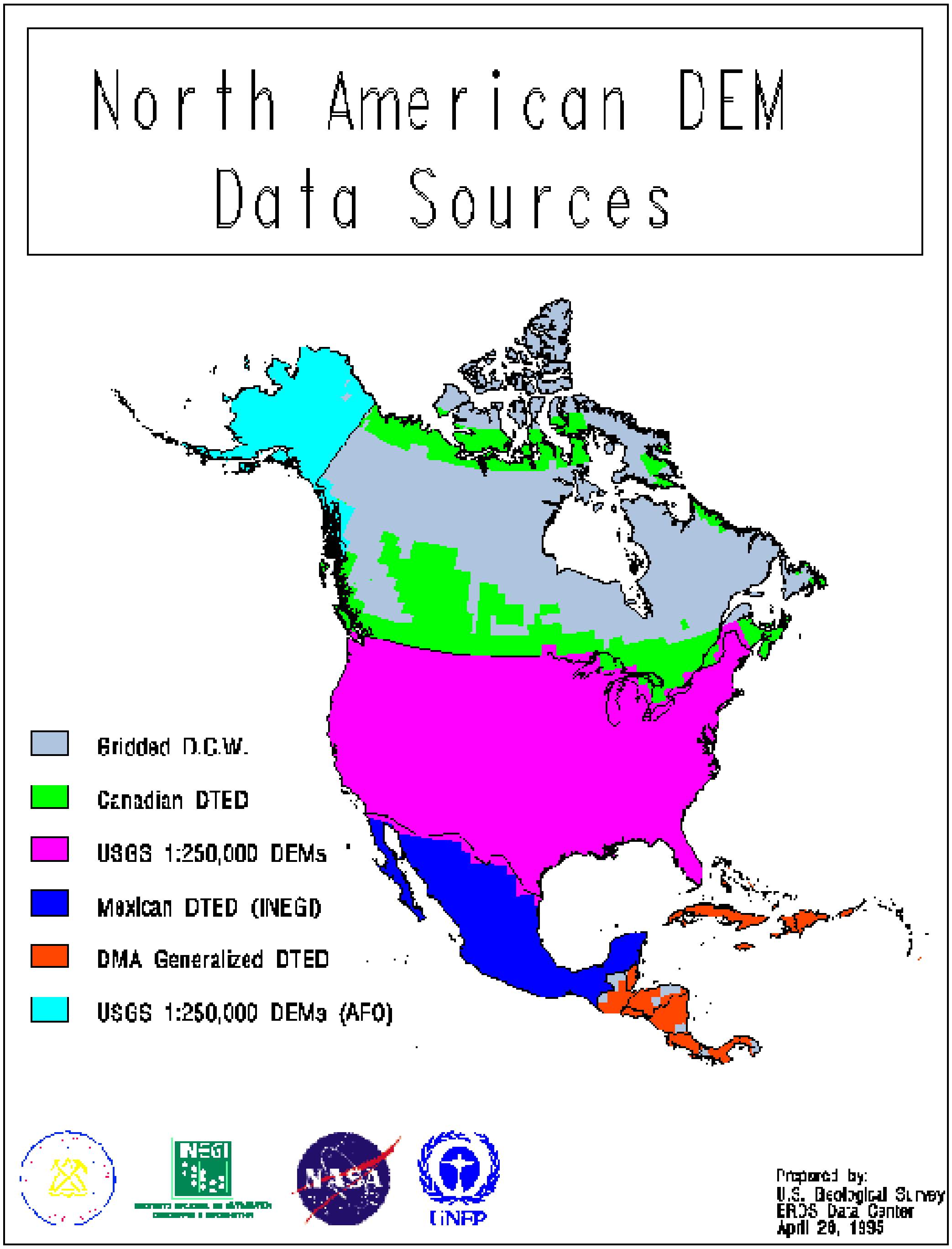
To reset this form, press the button - Reset **.**

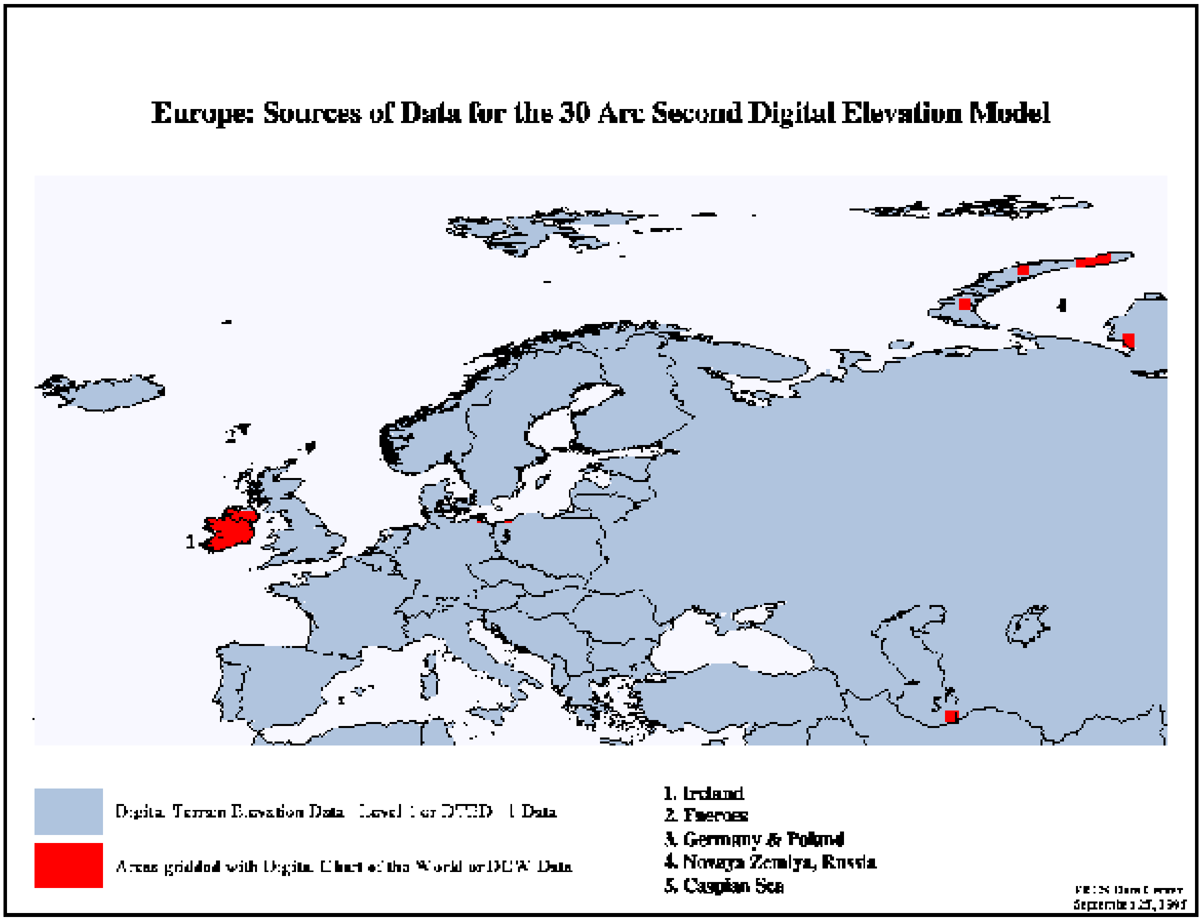

EIS Class Definition
Main server: shoofly.cs.umt.edu
Hierarchy: Landscape
Owner: (Not supported yet)
Object: hab_series
Class Name: hab_series
Parent Name: root
Description:
. . Interface Classes:
. . . . .NONE
. . Forward Declaration Classes:
. . . . .NONE
. . Attributes:
. . . . Name: series_code Type: hab_series_code
. . . . Name: series_scien Type: hab_series_scien
Options:
Close the current object
Close current hierarchy
Change the main server
Quit EIS
Get help
Act on Selected Option

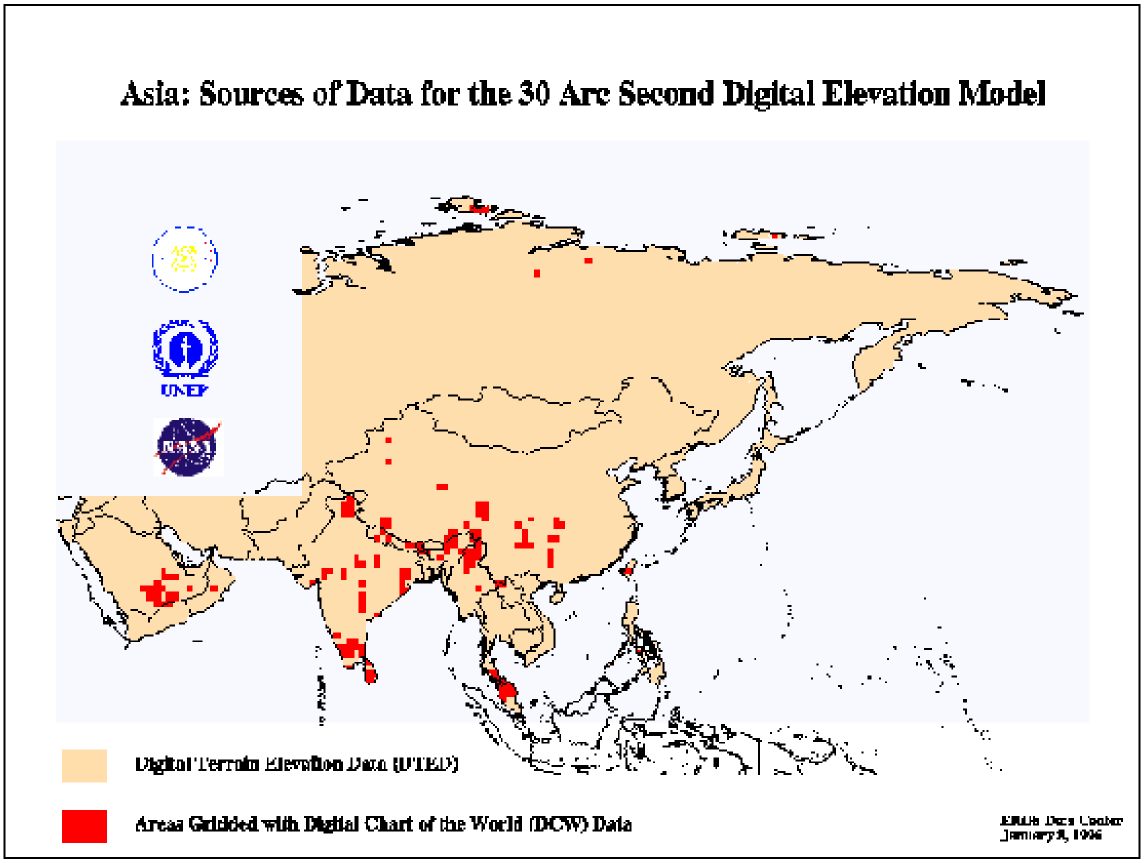

Create Class
Class Name
Description
(cannot be blank)
Class Components
Description
Class Interface
(optional) Explicit dependencies on other class definitions
Class Attributes
Visible types, variables, and operations
Class Parameter Declarations
(optional) Parameterized parts
Inherited Parameter Bindings
(optional) Bindings for parent class parameters
Documents
(optional) List of related documents
Create
Cancel
Close
Help

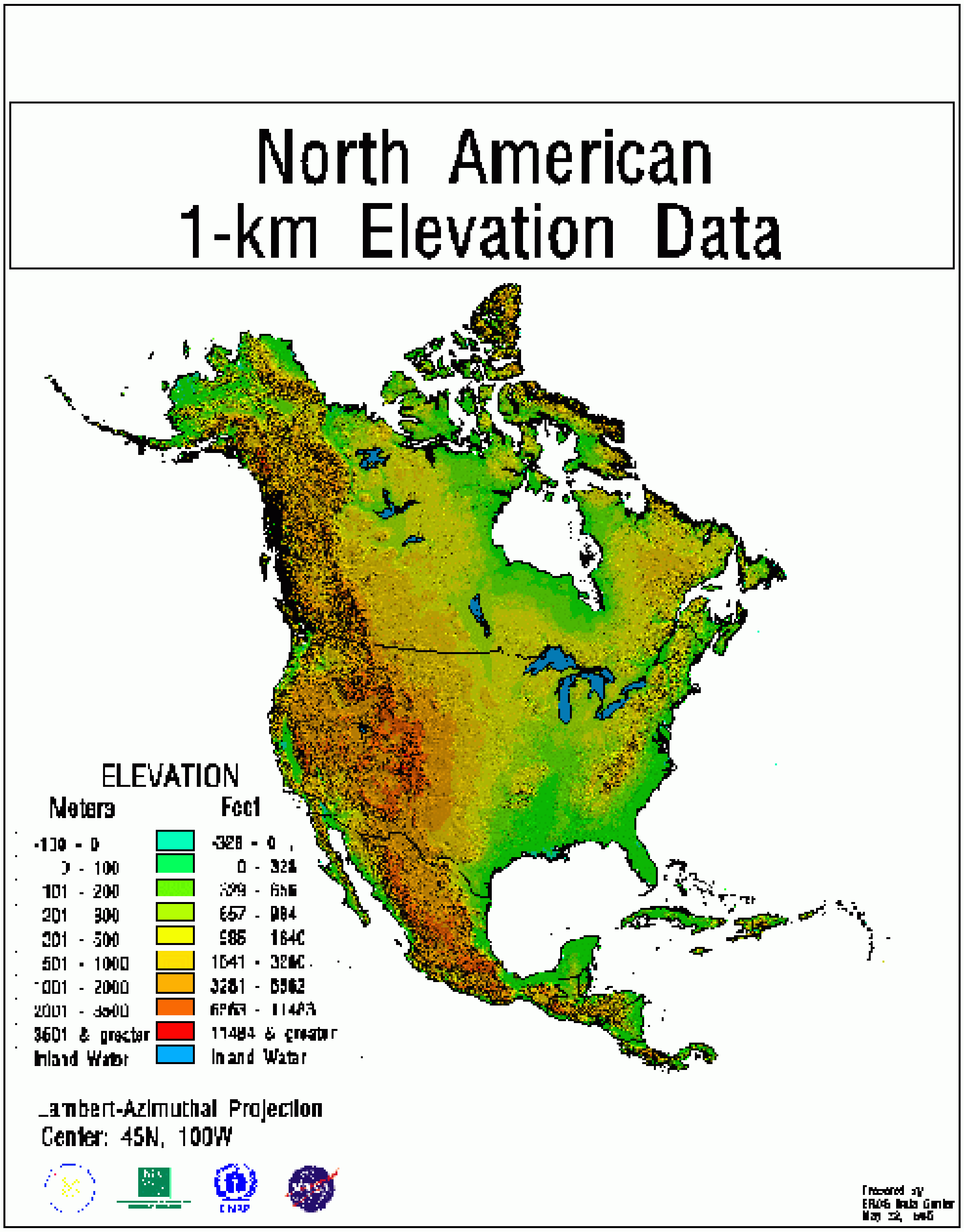

Ecosystem Information System - User Interface
File
Edit
Object
Tree
Groups
Permissions
Administration
Help
Discrete_Point_Model
Discrete_Line_Point_Model
elev_p39r28
Elevation
elev_p41r28
AI_DemLattice
asp_p39r28
Aspect
asp_p41r28
tm_p39r28
RawTM_MB_Model
Instances
Methods

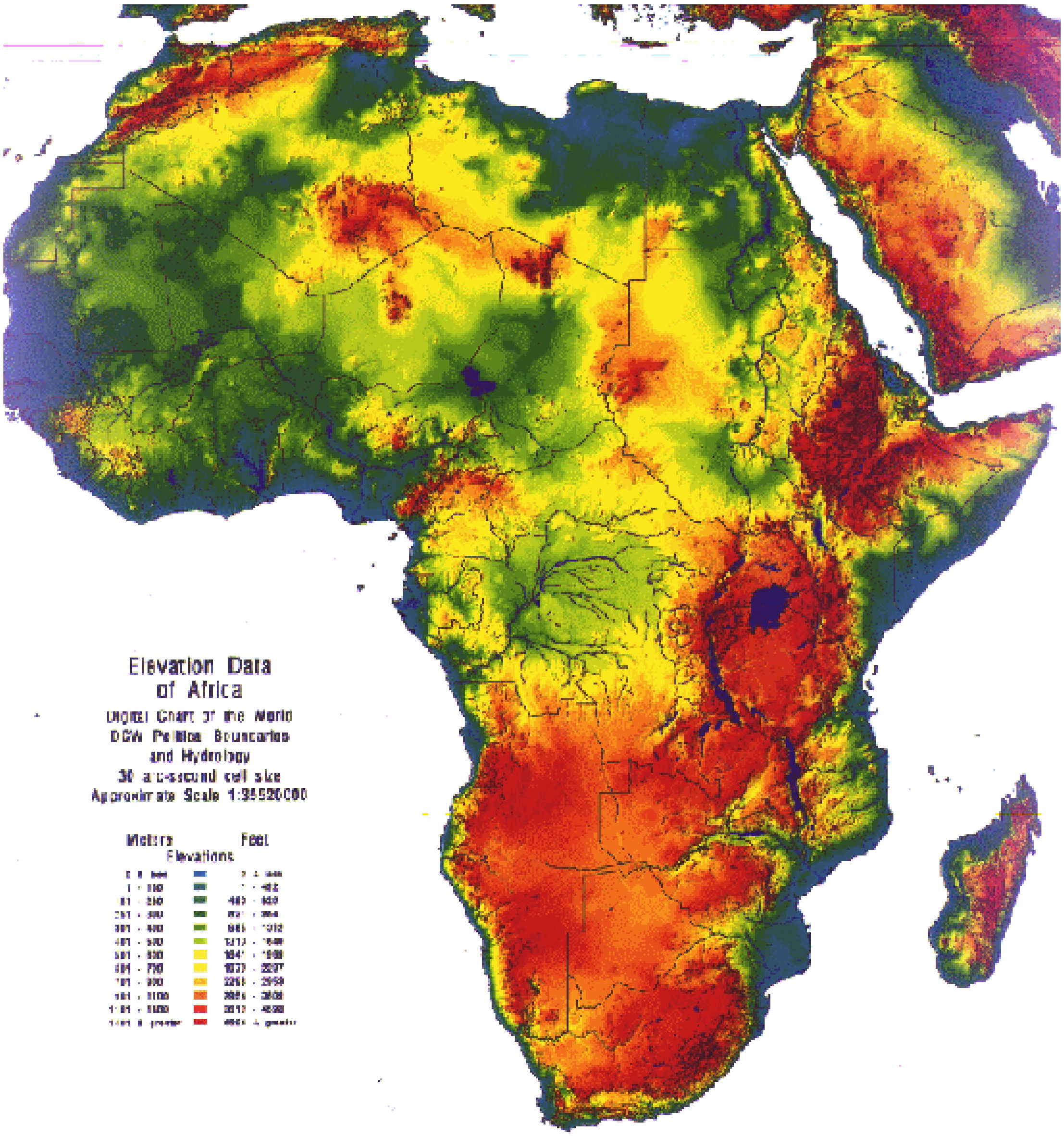

EIS Hierarchy Structure
You may select one of the objects in the hierarchy to open.
Main server: shoofly.cs.umt.edu
Hierarchy: r3_1
Owner: (Not supported yet)
Display Options:
Instances shown
Methods shown
Structure:
eisroot CLASS
. Data_Type CLASS
. . Simple_Data_Type CLASS
. . Complex_Data_Type CLASS
. . . Non_Id_Records CLASS
. . . . RHESSys_v1_Param CLASS
. . . . ERDAS_HDR CLASS
. . . . ERDAS_STATS CLASS
. . . . ERDAS_TRAIL CLASS
Options:

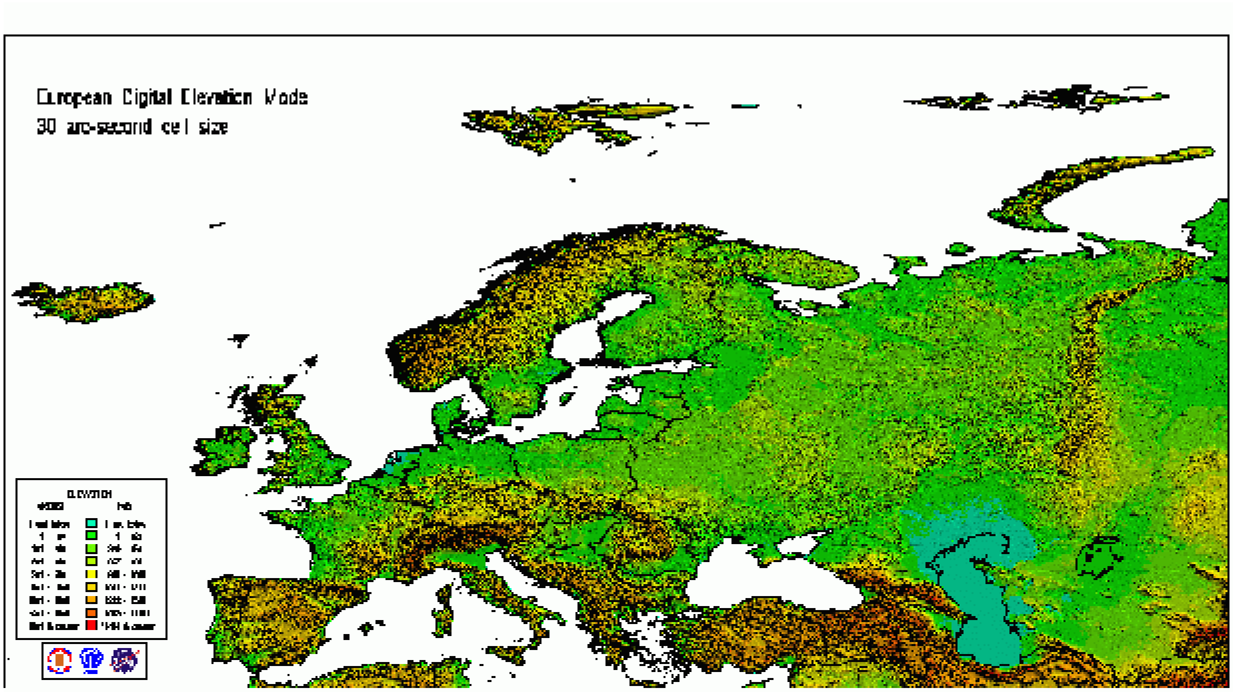

North American DEM
Data Sources
Gridded D.C.W.
Canadian DTED
USGS 1:250,000 DEMs
Mexican DTED (INEGI)
DMA Generalized DTED
USGS 1:250,000 DEMs (AFO)
NASA
UNEP
Prepared by:
U.S. Geological Survey
EROS Data Center
April 20, 1995

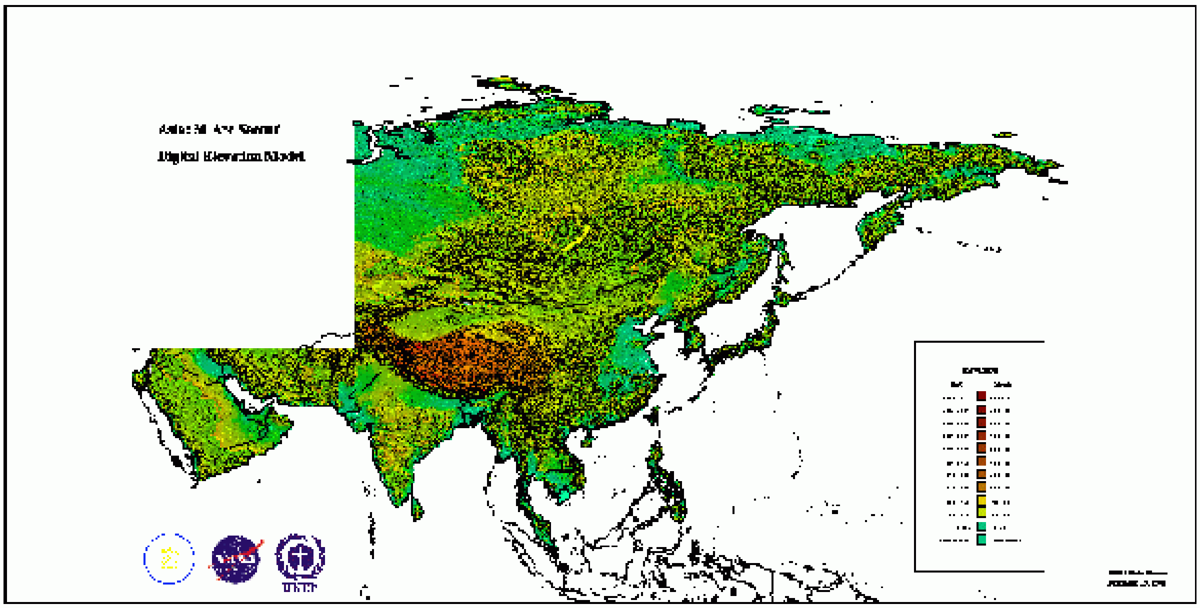

Europe: Sources of Data for the 30 Arc Second Digital Elevation Model
1
2
3
4
5
Digital Terrain Elevation Data – Level 1 or DTED – 1 Data
Areas gridded with Digital Chart of the World or DCW Data
1. Ireland
2. Faeroes
3. Germany & Poland
4. Novaya Zemlya, Russia
5. Caspian Sea

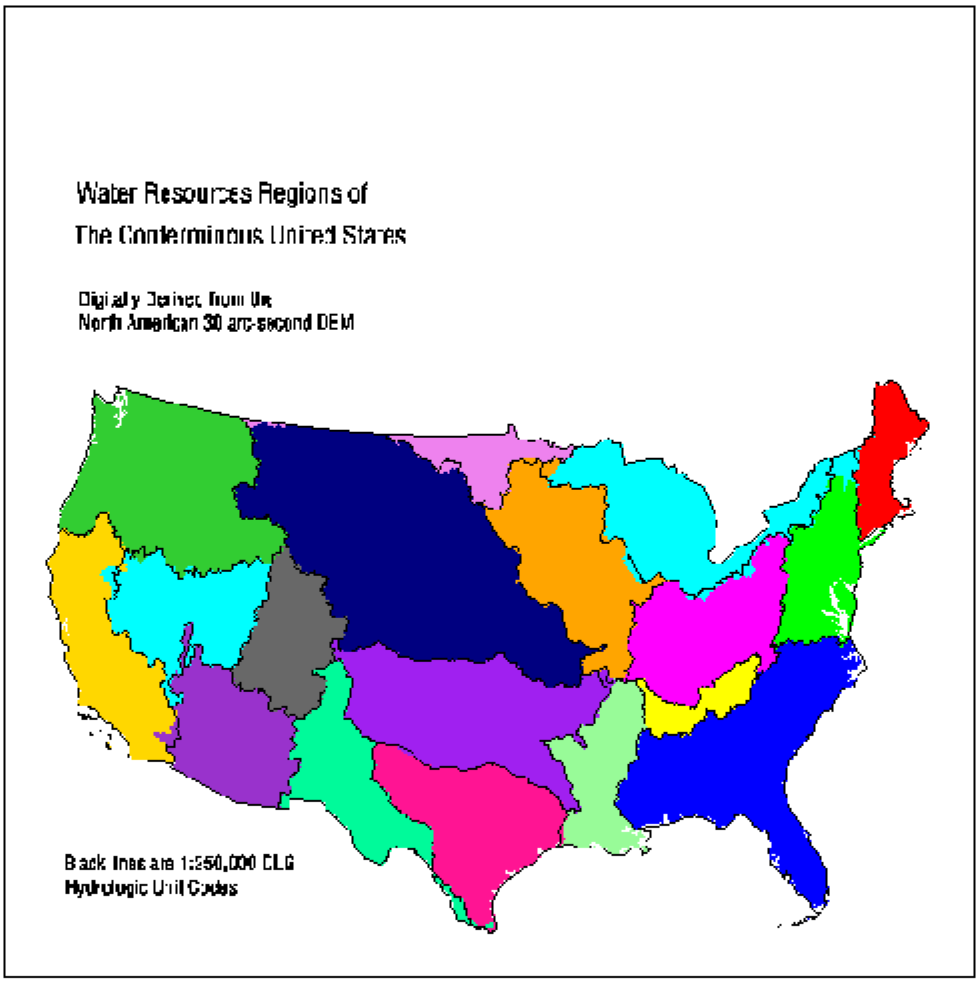

Asia: Sources of Data for the 30 Arc Second Digital Elevation Model
UNEP
Digital Terrain Elevation Data (DTED)
Areas Gridded with Digital Chart of the World (DCW) Data


North American
1-km Elevation Data
ELEVATION
Meters
Feet
0 - 328
101 - 200
657 - 984
501 - 1000
1041 - 3280
1001 - 2000
3281 - 6562
Inland Water
Lambert-Azimuthal Projection
Center: 45N, 100W

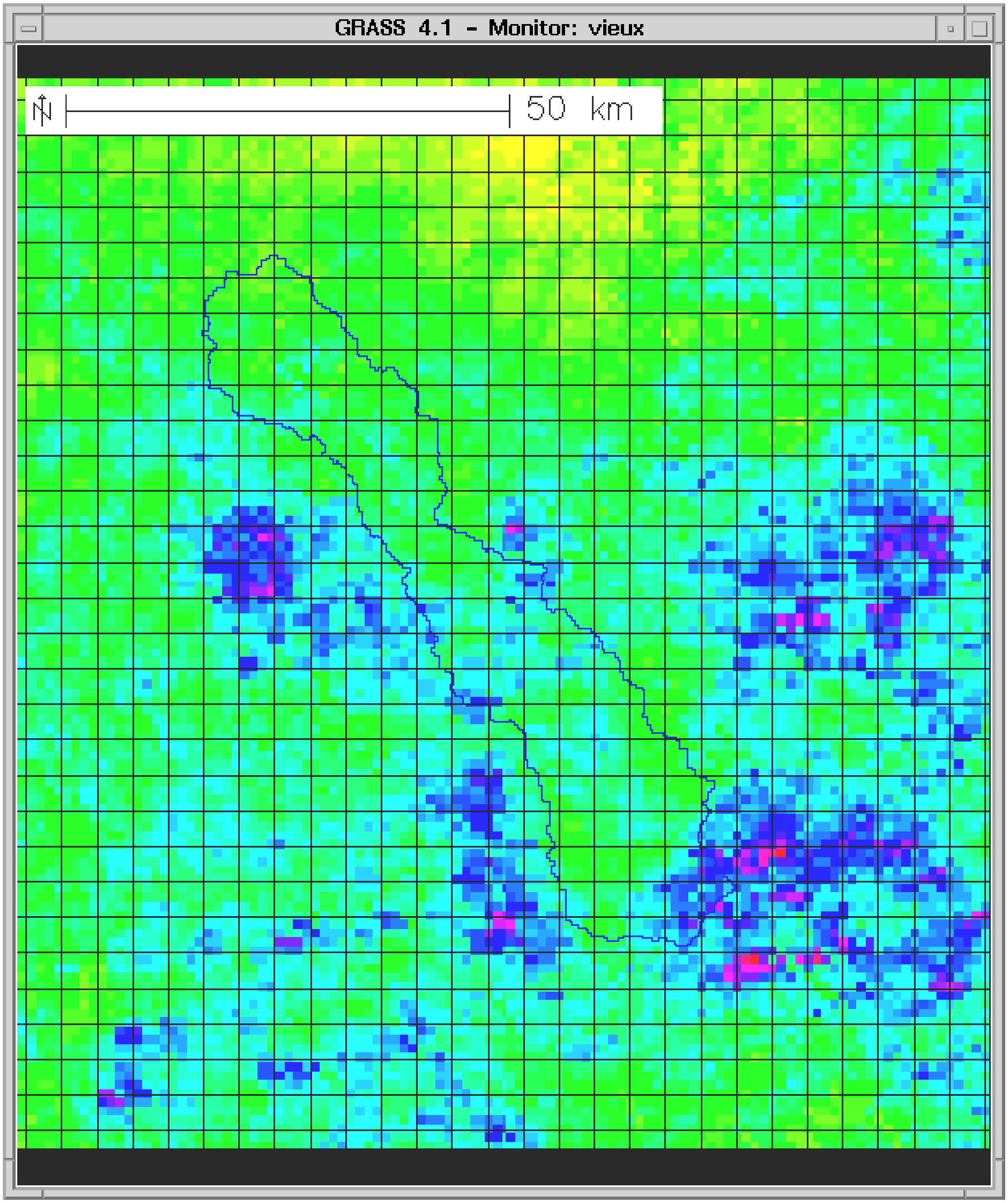

Elevation Data
of Africa
and Hydrology
Meters
Feet
Elevations

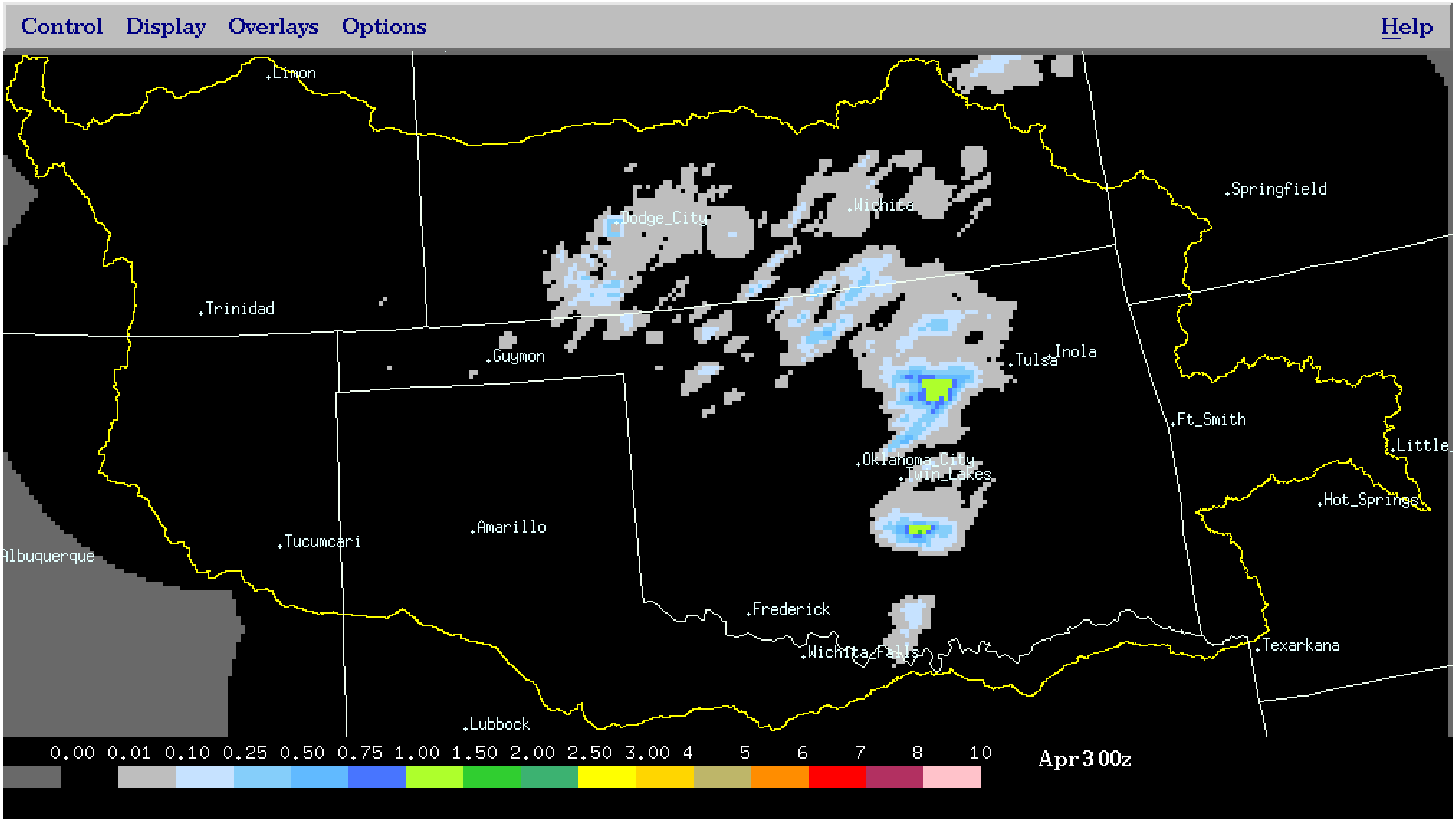

European Digital Elevation Model
30 arc-second cell size

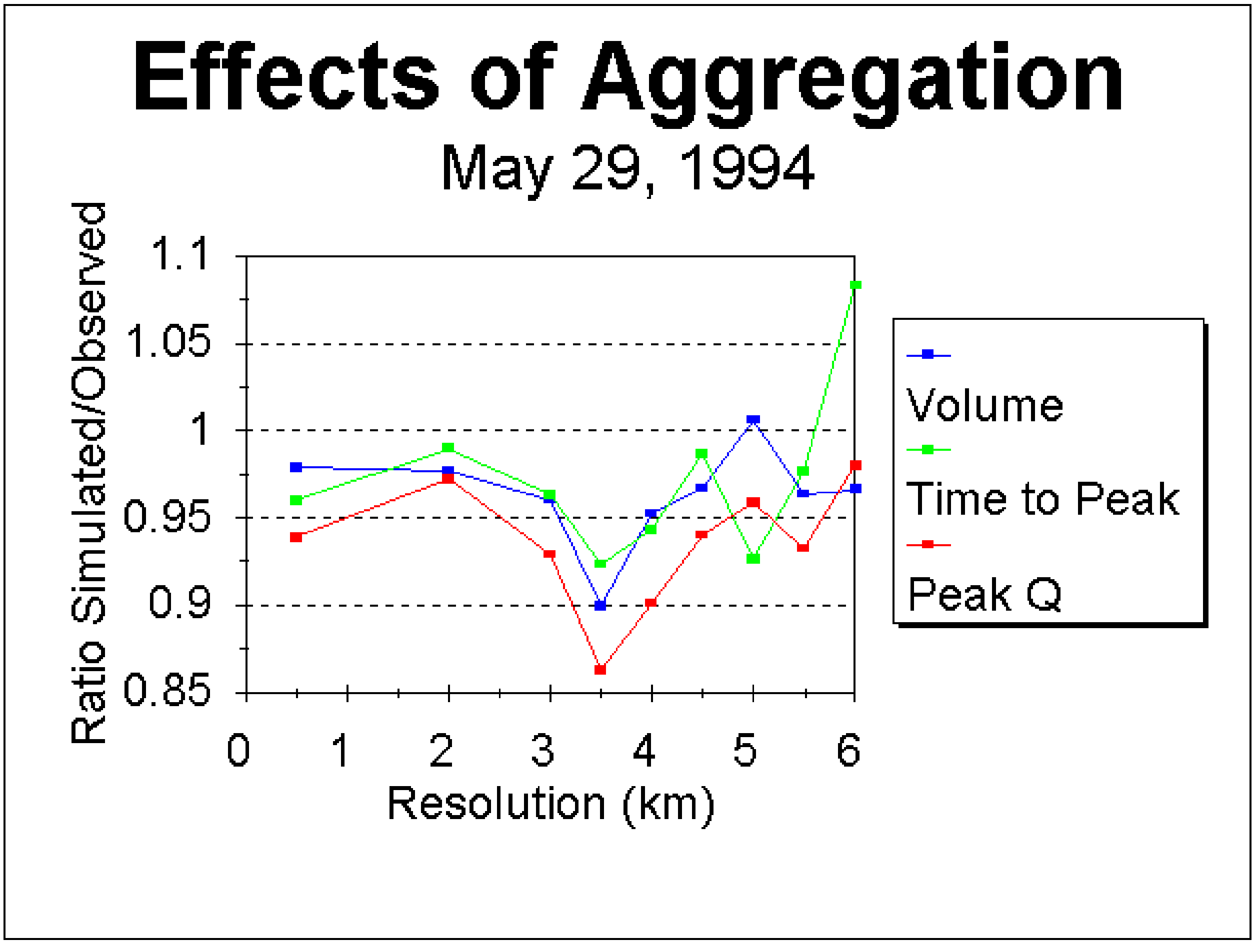

Digital Elevation Model


Water Resources Regions of
The Conterminous United States
North American 30 arc-second DEM
Black lines are 1:250,000 DLG
Hydrologic Unit Codes


Discrepancies between USGS HUCs
And Regions Derived from 30 arc-second DEM


GRASS 4.1 - Monitor: vieux
N
50 km


Control
Display
Overlays
Options
Help
Limon
Springfield
Wichita
Dodge_City
Trinidad
Guymon
Tulsa
Inola
Ft_Smith
Little
Oklahoma_City
Twin_Lakes
Hot_Springs
Amarillo
Tucumcari
Albuquerque
Frederick
Wichita_Falls
Texarkana
Lubbock
0.00 0.01 0.10 0.25 0.50 0.75 1.00 1.50 2.00 2.50 3.00 4 5 6 7 8 10
Apr 3 00z


Effects of Aggregation
May 29, 1994
Ratio Simulated/Observed
1.1
1.05
1
0.95
0.9
0.85
0
1
2
3
4
5
6
Resolution (km)
Volume
Time to Peak
Peak Q

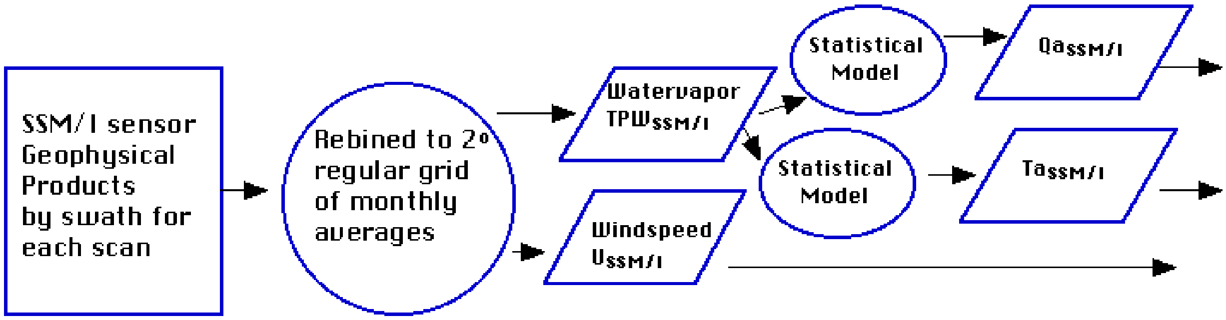

SSM/I sensor Geophysical Products by swath for each scan
Rebined to 2° regular grid of monthly averages
Watervapor TPWSSM/I
Windspeed USSM/I
Statistical Model
Statistical Model
QaSSM/I
TaSSM/I

Figure 1

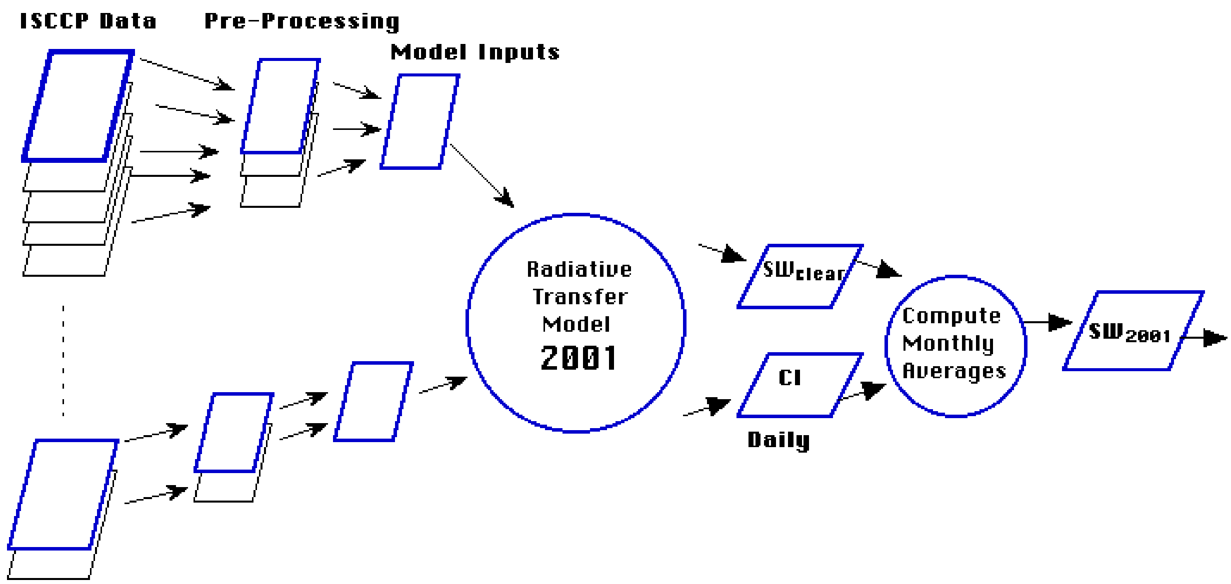
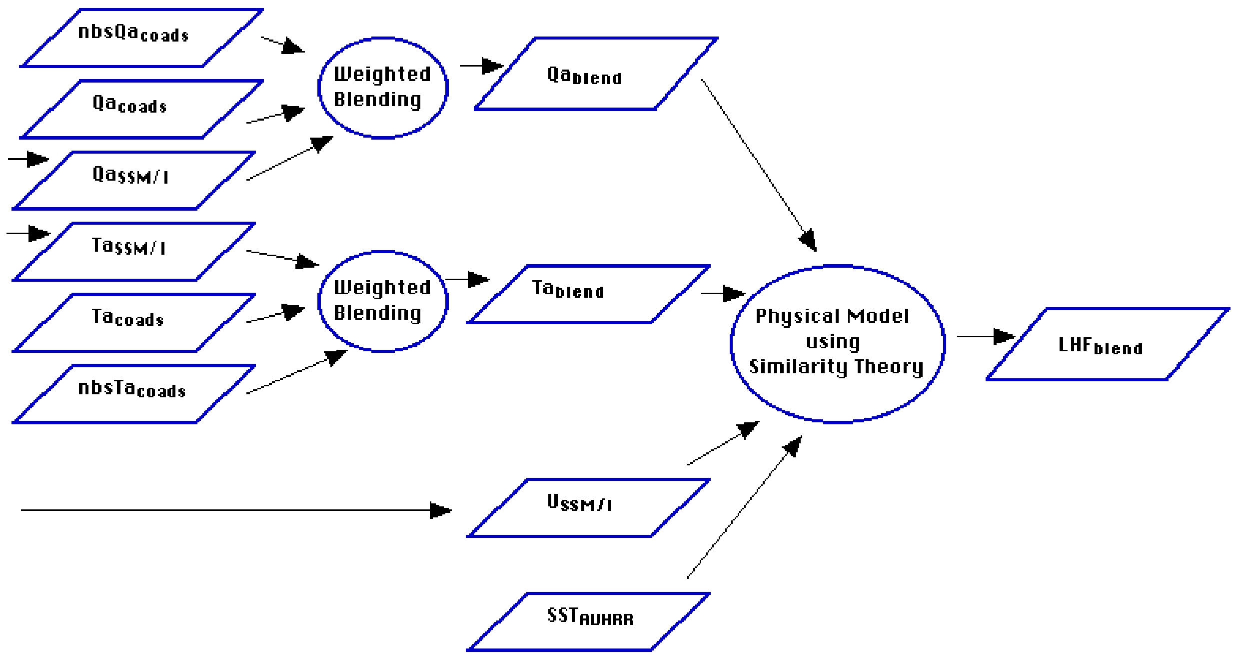

ISCCP Data
Pre-Processing
Model Inputs
Radiative Transfer Model 2001
SWclear
CI
Daily
Compute Monthly Averages
SW2001

Figure 2



nbsQa_coads
Qa_coads
Qa_SSM/I
Ta_SSM/I
Ta_coads
nbsTa_coads
Weighted Blending
Weighted Blending
Qa_blend
Ta_blend
U_SSM/I
SST_AVHRR
Physical Model using Similarity Theory
LHF_blend

Figure 3

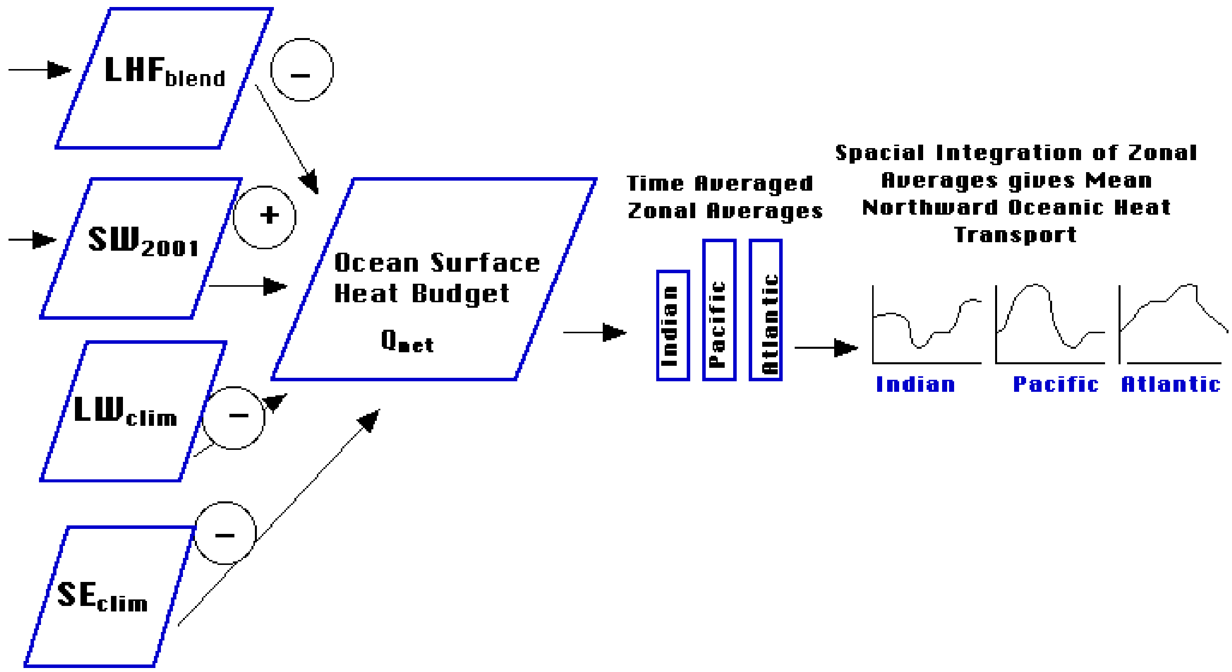
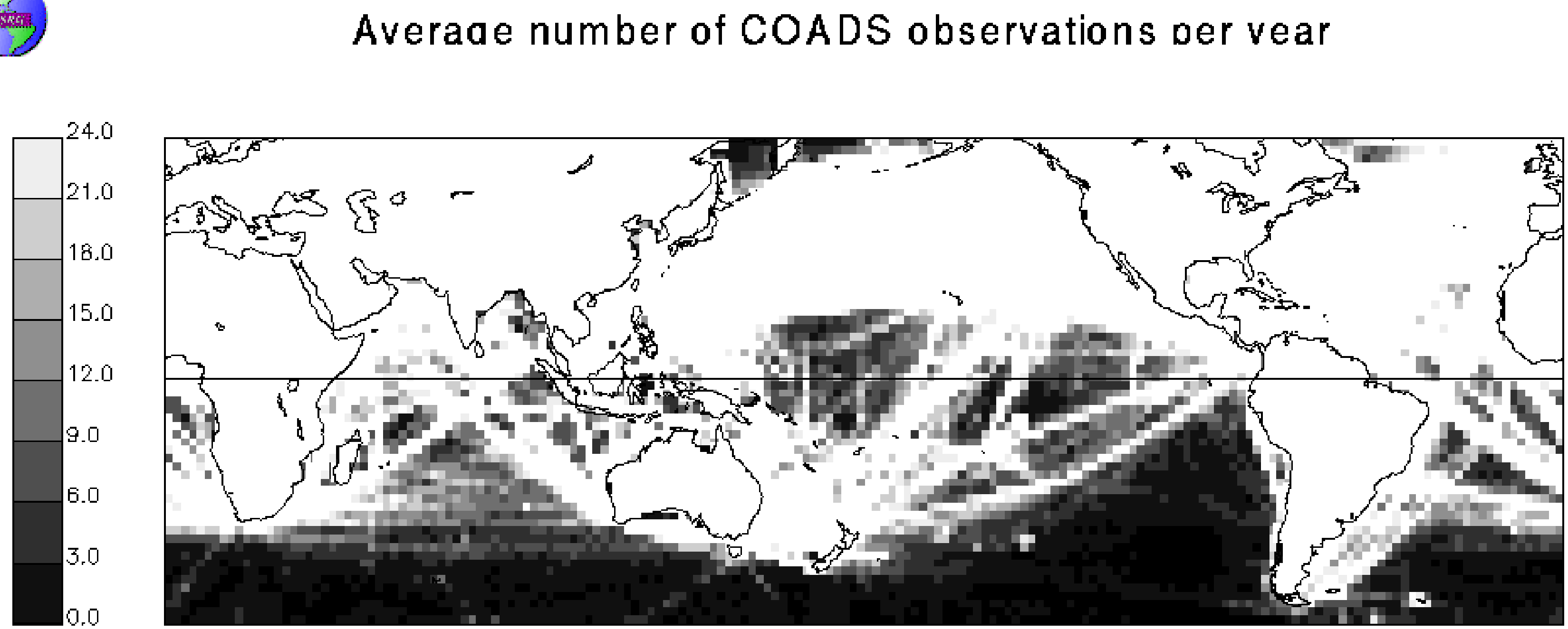
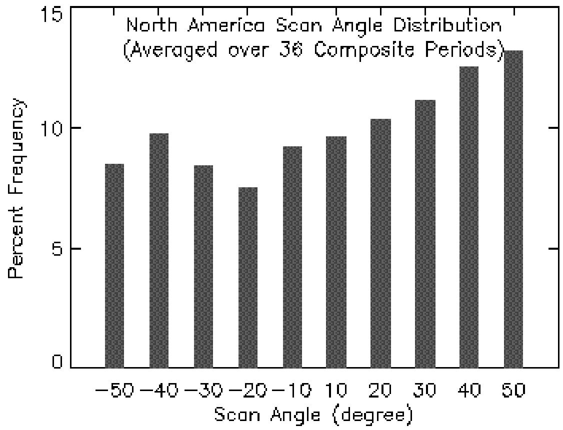
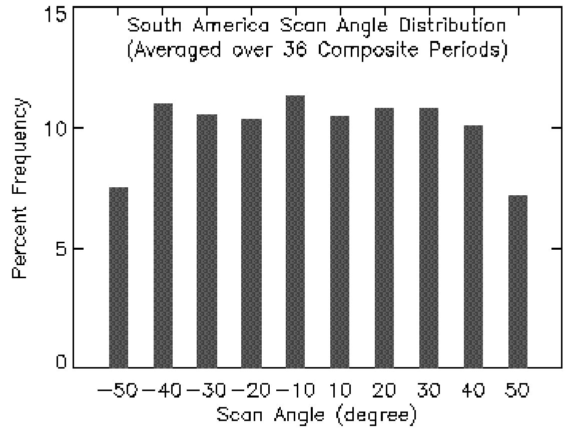

LHF blend
SW 2001
LW clim
SE clim
−
+
−
−
Ocean Surface
Heat Budget
Q net
Time Averaged
Zonal Averages
Indian
Pacific
Atlantic
Spacial Integration of Zonal
Averages gives Mean
Northward Oceanic Heat
Transport
Indian
Pacific
Atlantic

Figure 4

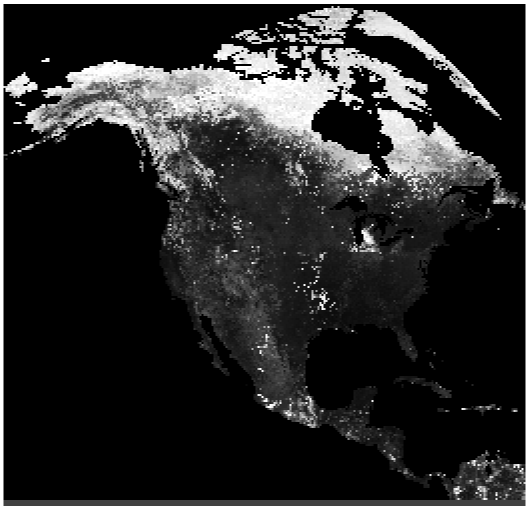

Average number of COADS observations per year
24.0
21.0
18.0
15.0
12.0
9.0
6.0
3.0
0.0


North America Scan Angle Distribution
(Averaged over 36 Composite Periods)
Percent Frequency
15
10
5
0
−50 −40 −30 −20 −10 10 20 30 40 50
Scan Angle (degree)


South America Scan Angle Distribution
(Averaged over 36 Composite Periods)
Percent Frequency
15
10
5
0
−50 −40 −30 −20 −10 10 20 30 40 50
Scan Angle (degree)

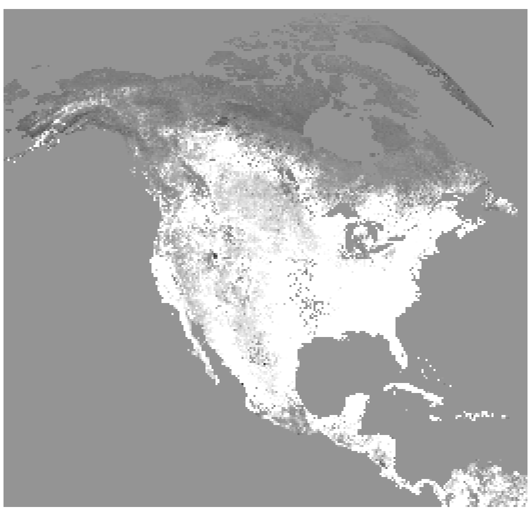
This file is not available. If you wish to obtain a copy of this file, please contact Limin Yang directly at:

Telephone: (605) 594-6039
Fax: (605) 594-6589
Email: lyang@unl.cr.usgs.gov

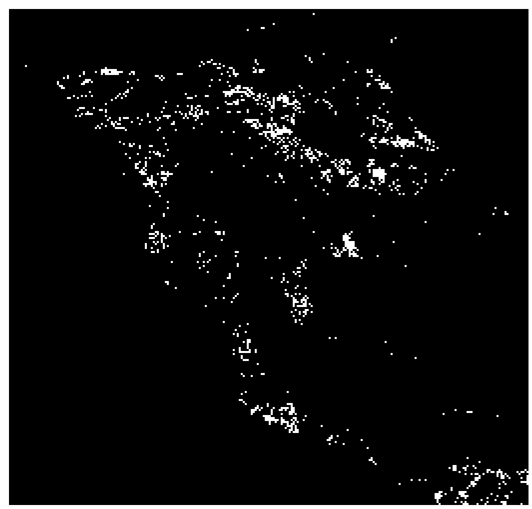


AVHRR band 1

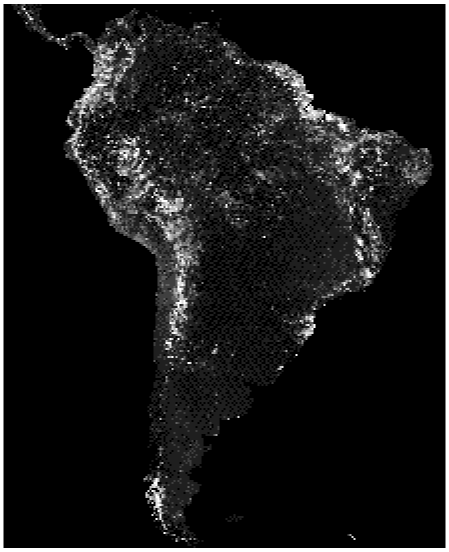
AVHRR band 4



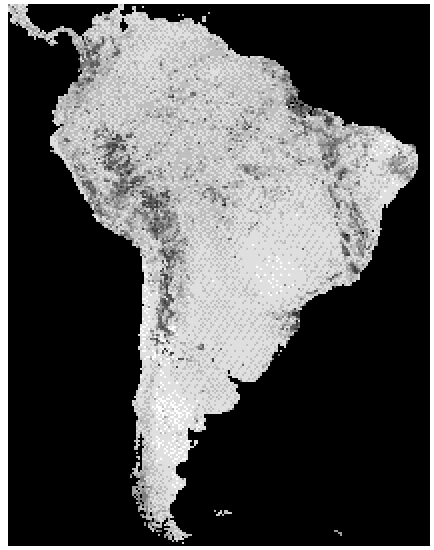
NDVI



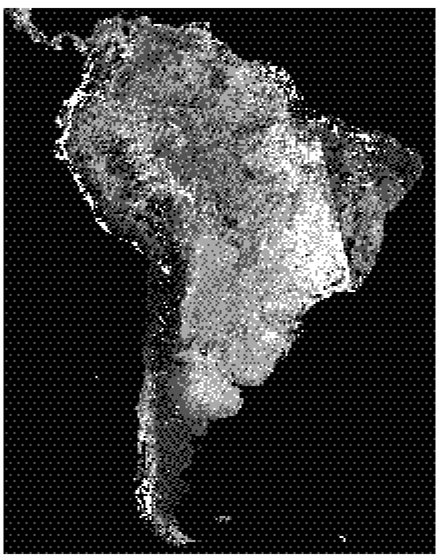
Cloud



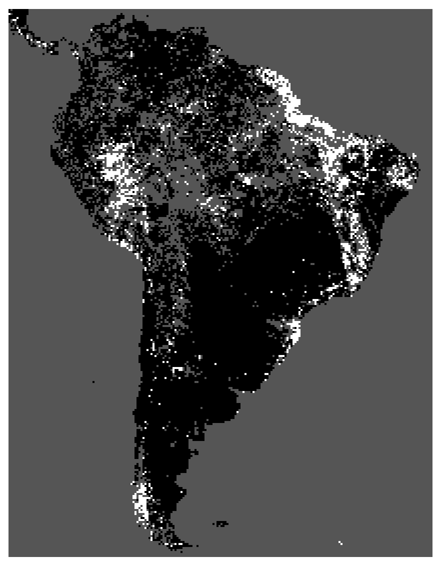
AVHRR band 1



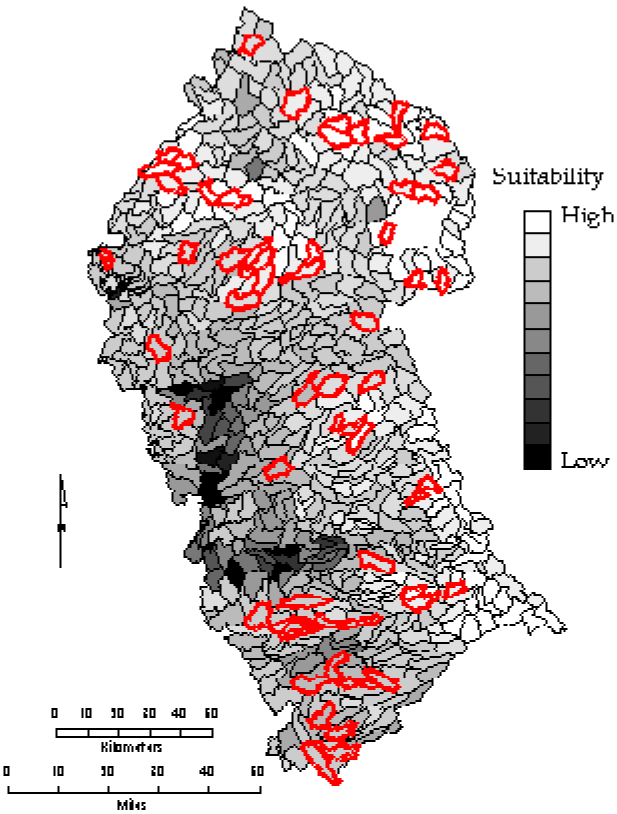
AVHRR band 4

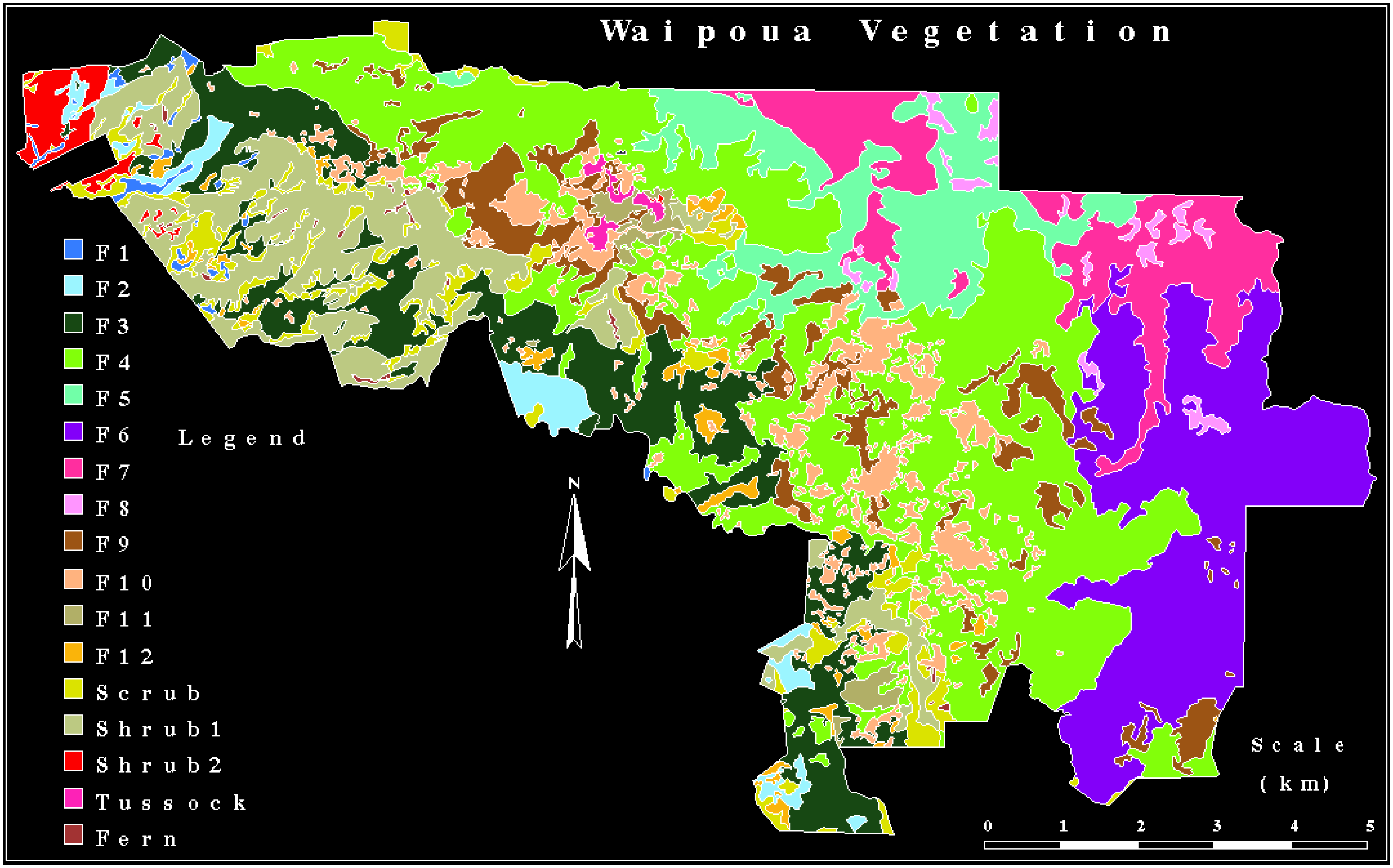


NDVI



cloud


Suitability
High
Low
Kilometers
Miles


Waipoua Vegetation
Legend
F 1
F 2
F 3
F 4
F 5
F 6
F 7
F 8
F 9
F 10
F 11
F 12
Scrub
Shrub1
Shrub2
Tussock
Fern
N
Scale
(km)
0
1
2
3
4
5

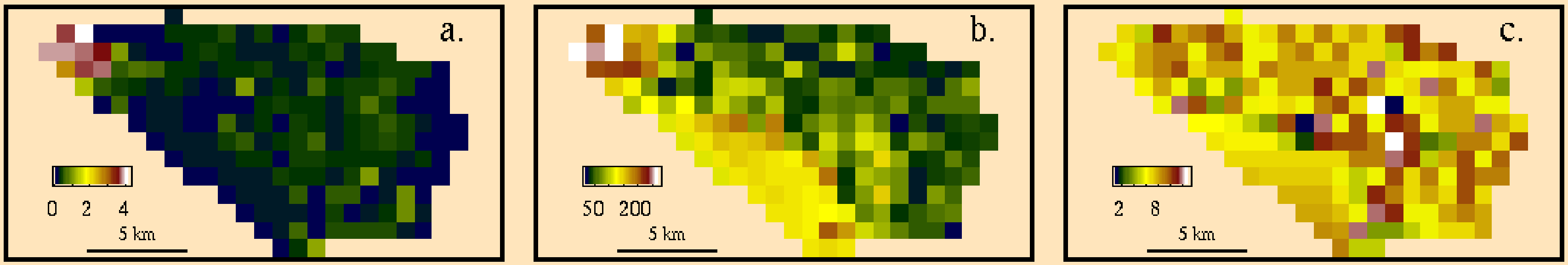


Figure 2.

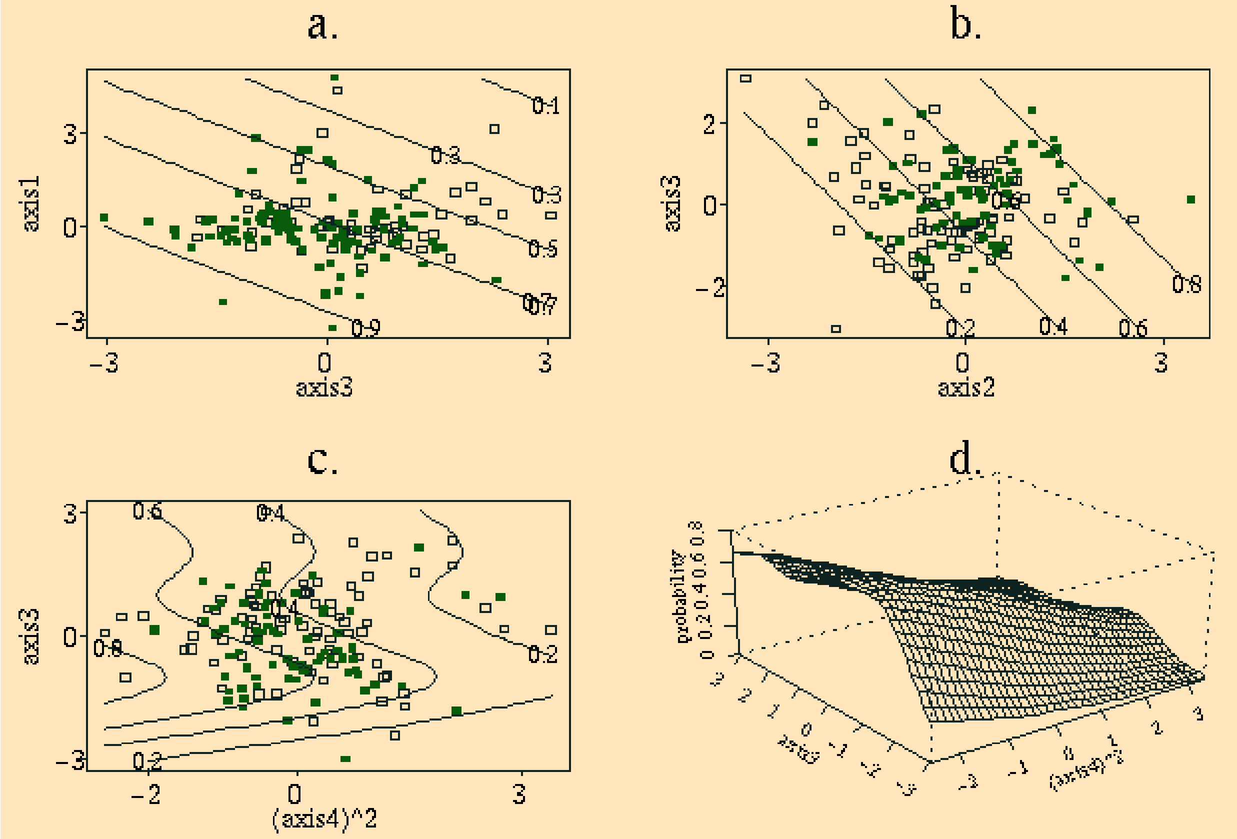


Figure 3.

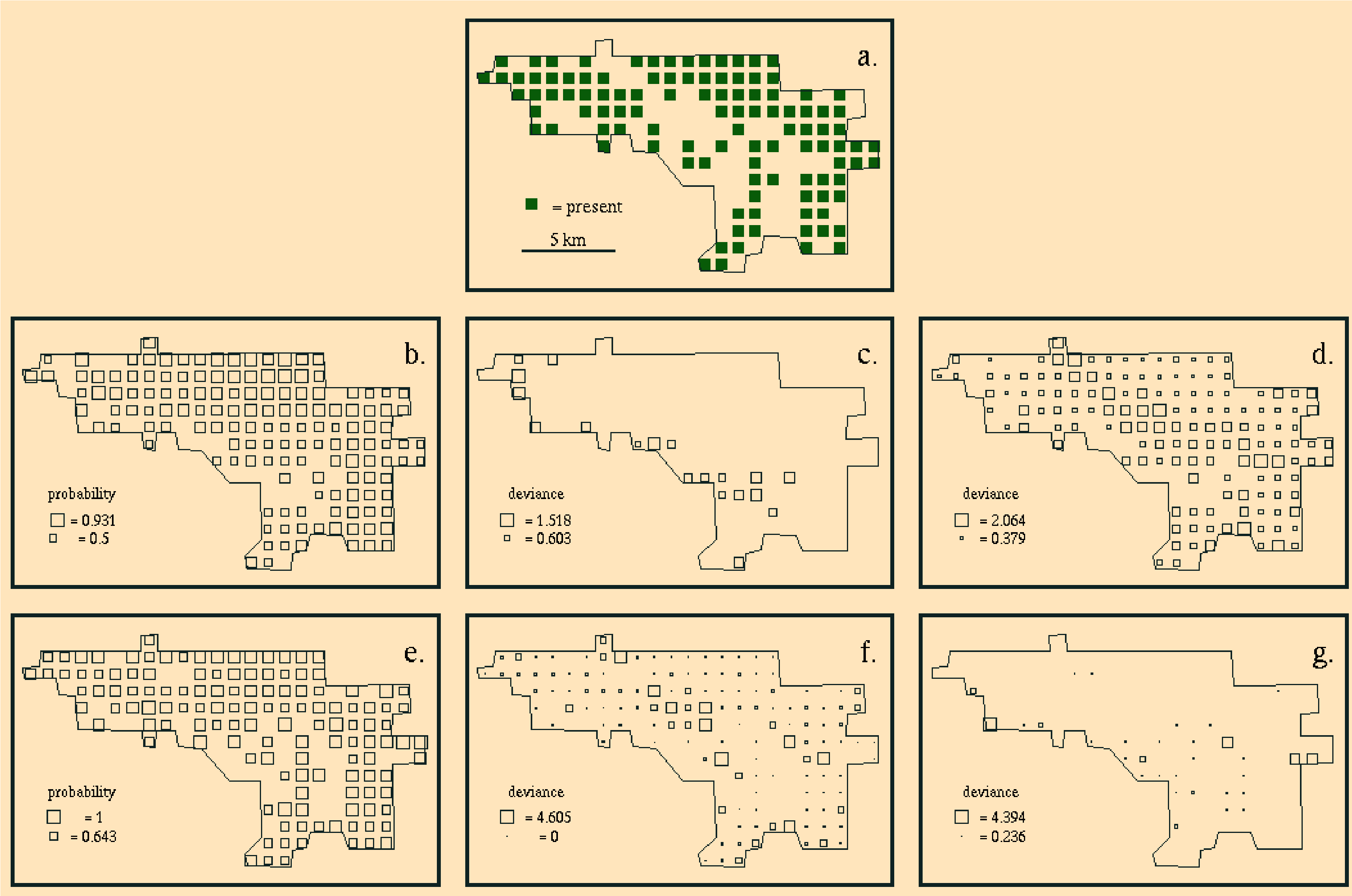

a.
= present
5 km
b.
probability
= 0.931
= 0.5
c.
deviance
= 1.518
= 0.603
d.
deviance
= 2.064
= 0.379
e.
probability
= 1
= 0.643
f.
deviance
= 4.605
= 0
g.
deviance
= 4.394
= 0.236

Figure 4.

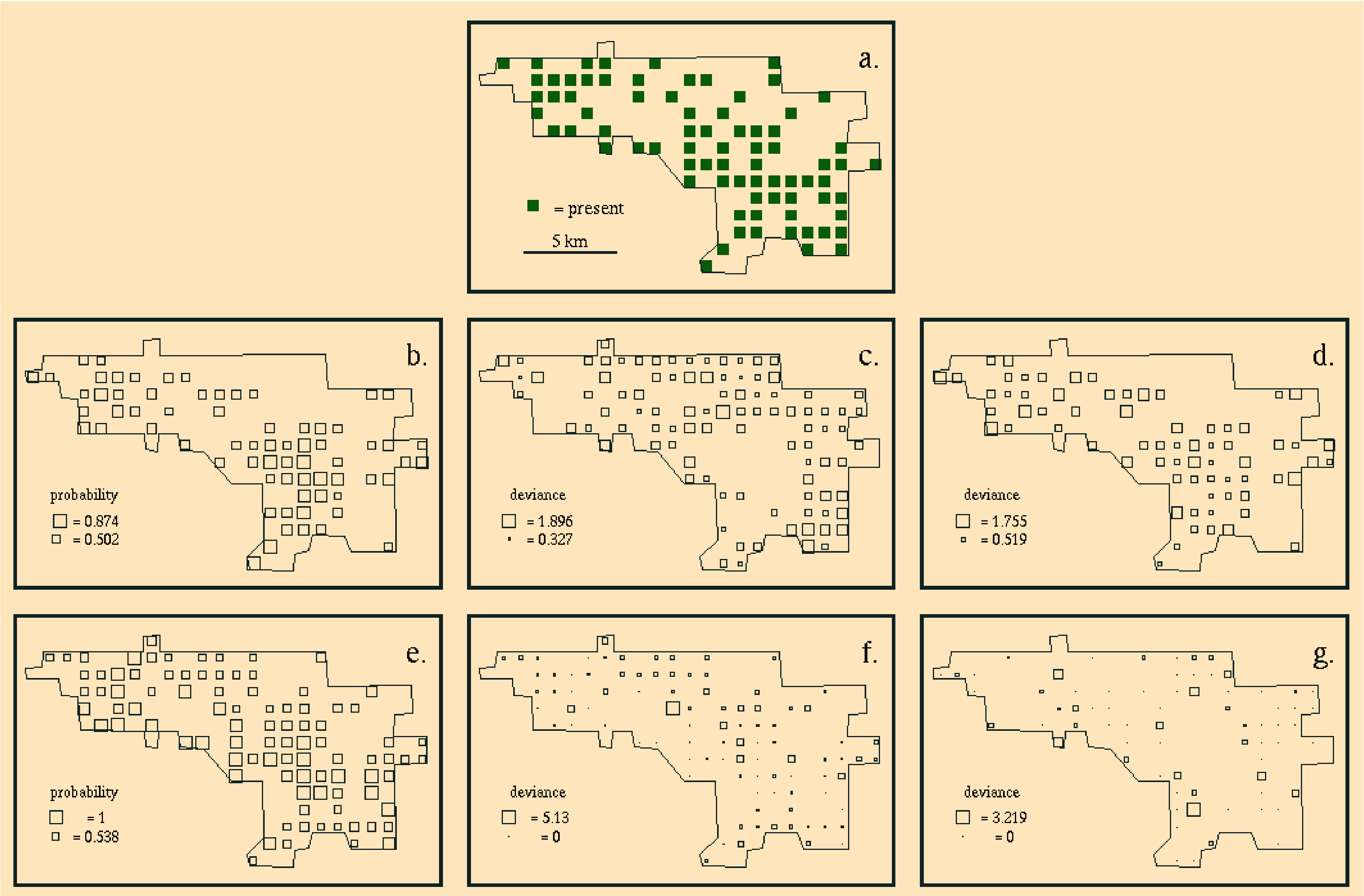
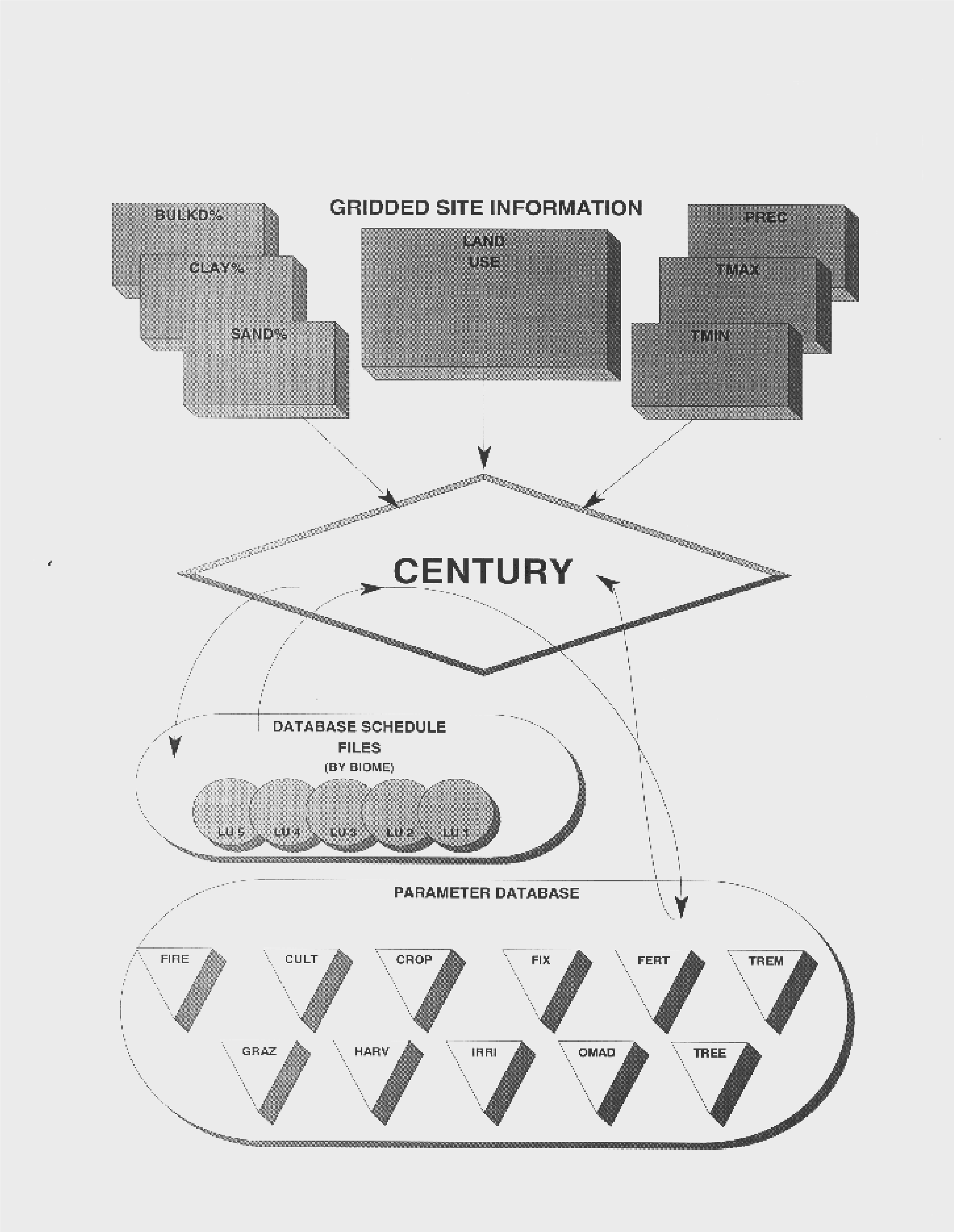
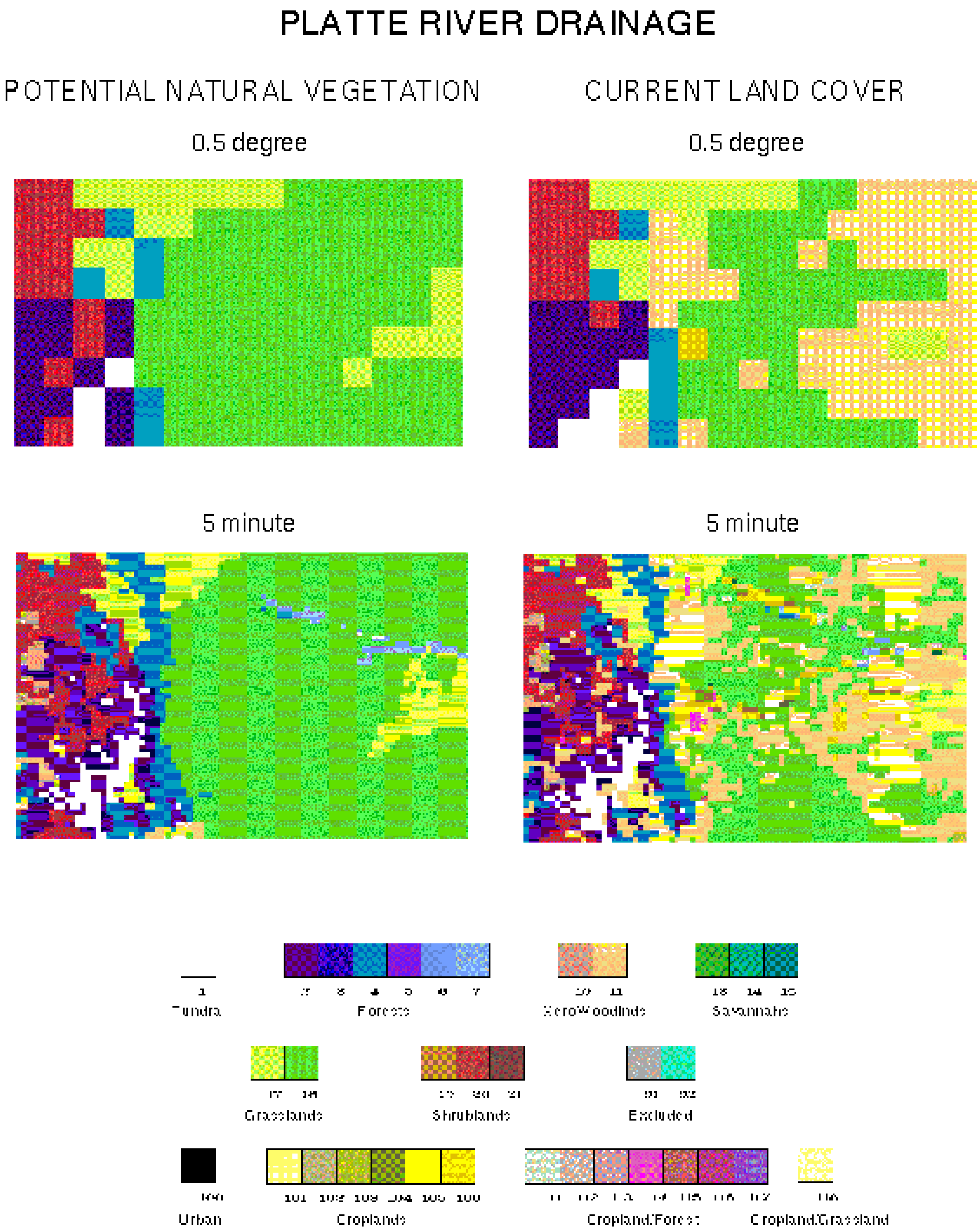
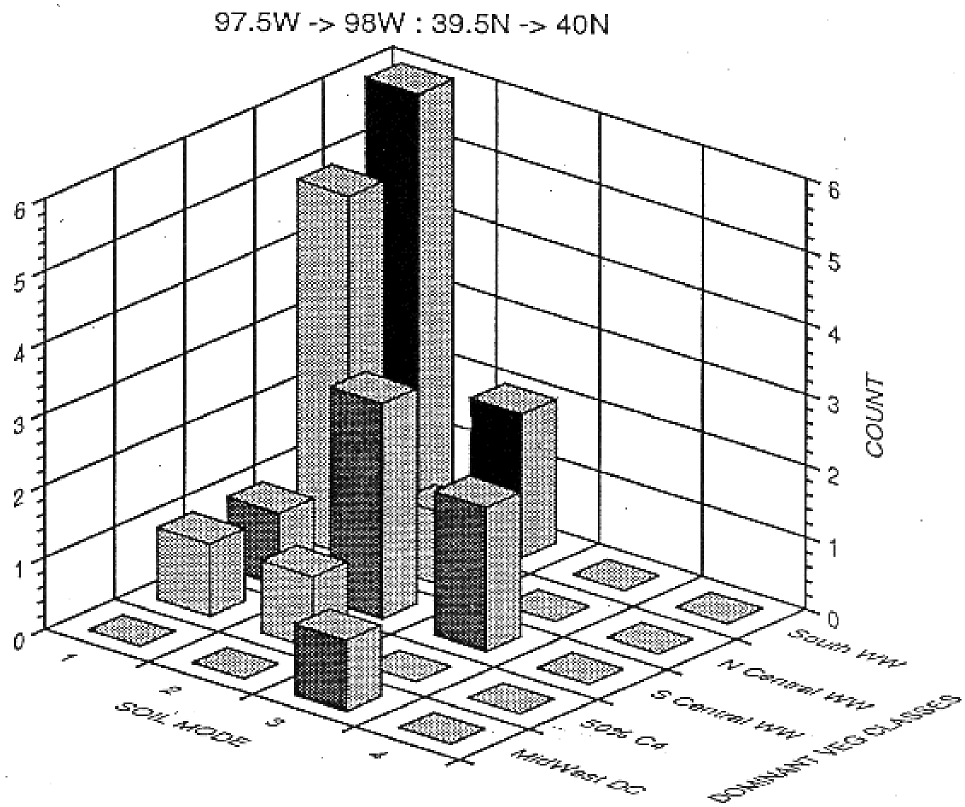
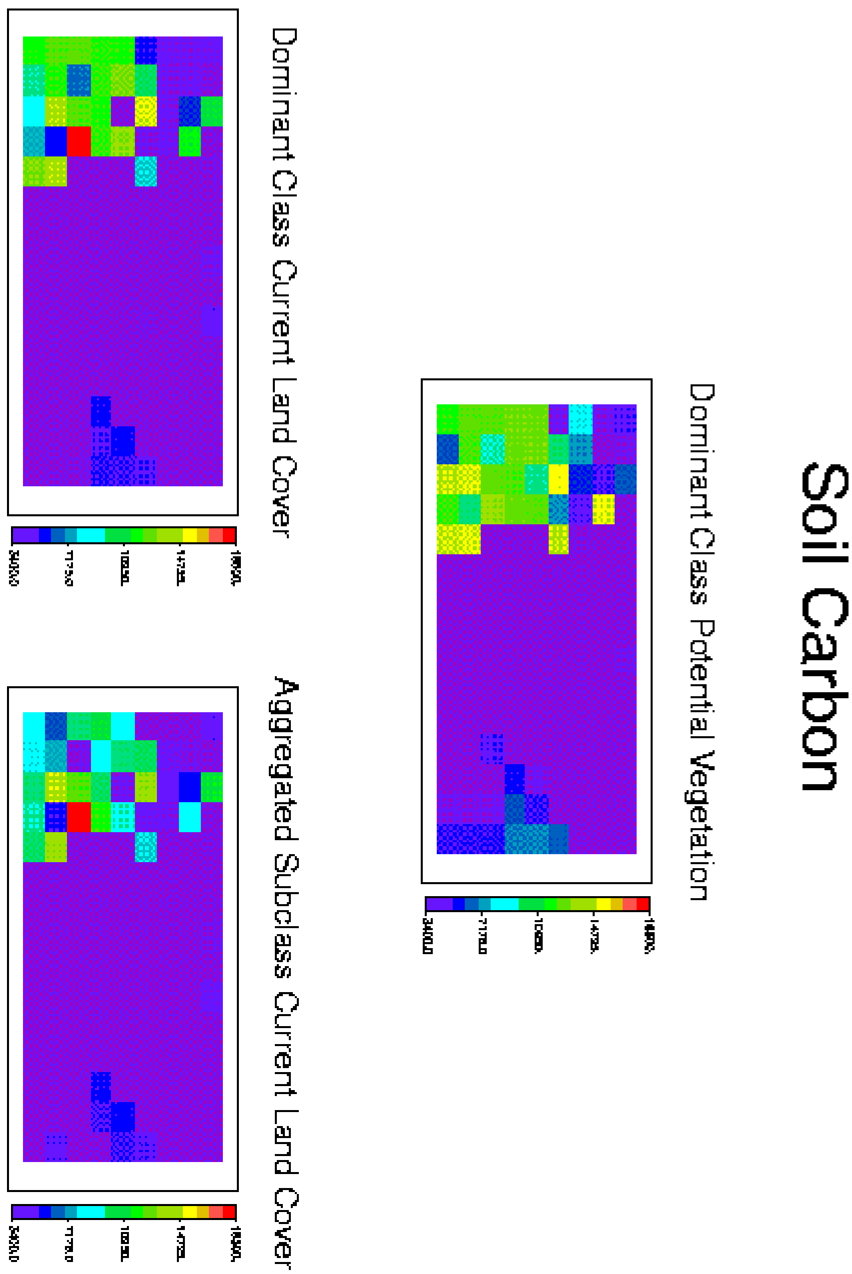

a.
= present
5 km
b.
probability
= 0.874
= 0.502
c.
deviance
= 1.896
= 0.327
d.
deviance
= 1.755
= 0.519
e.
probability
= 1
= 0.538
f.
deviance
= 5.13
= 0
g.
deviance
= 3.219
= 0

Figure 5.

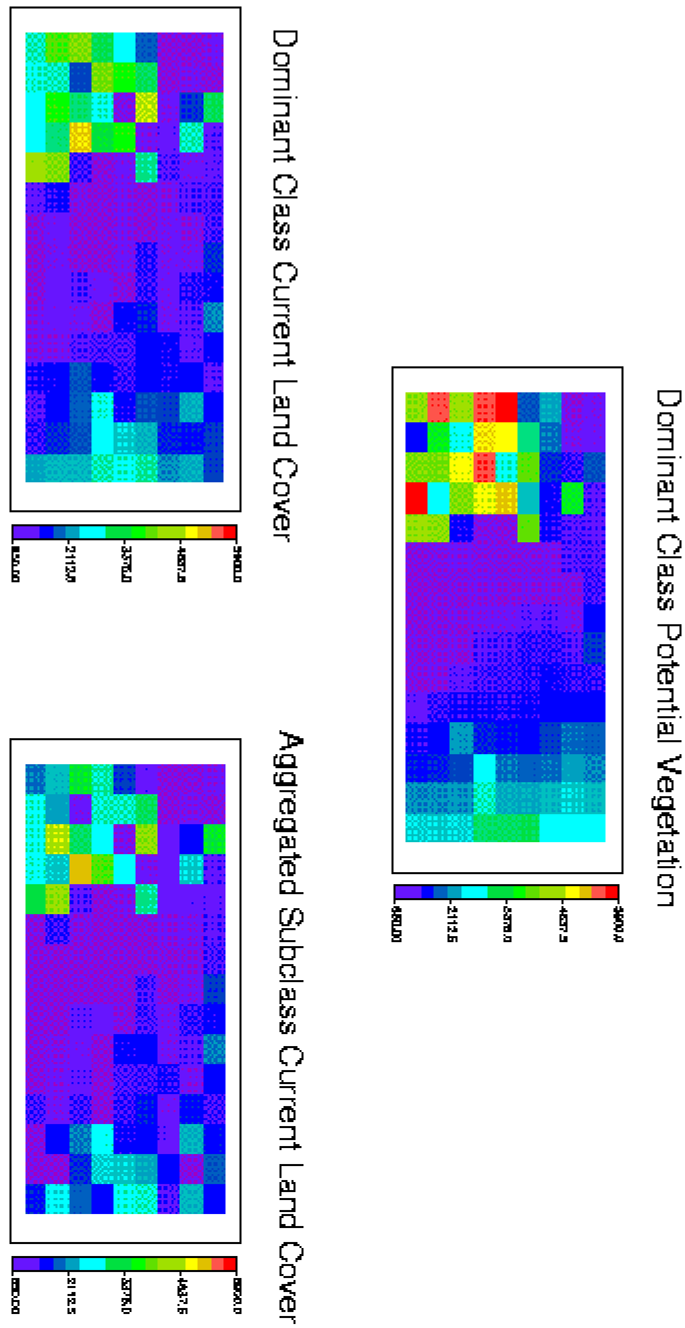

GRIDDED SITE INFORMATION
BULKD%
CLAY%
SAND%
LAND USE
PREC
TMAX
TMIN
CENTURY
DATABASE SCHEDULE FILES
(BY BIOME)
LU 5
LU 4
LU 3
LU 2
LU 1
PARAMETER DATABASE
FIRE
CULT
CROP
FIX
FERT
TREM
GRAZ
HARV
IRRI
OMAD
TREE

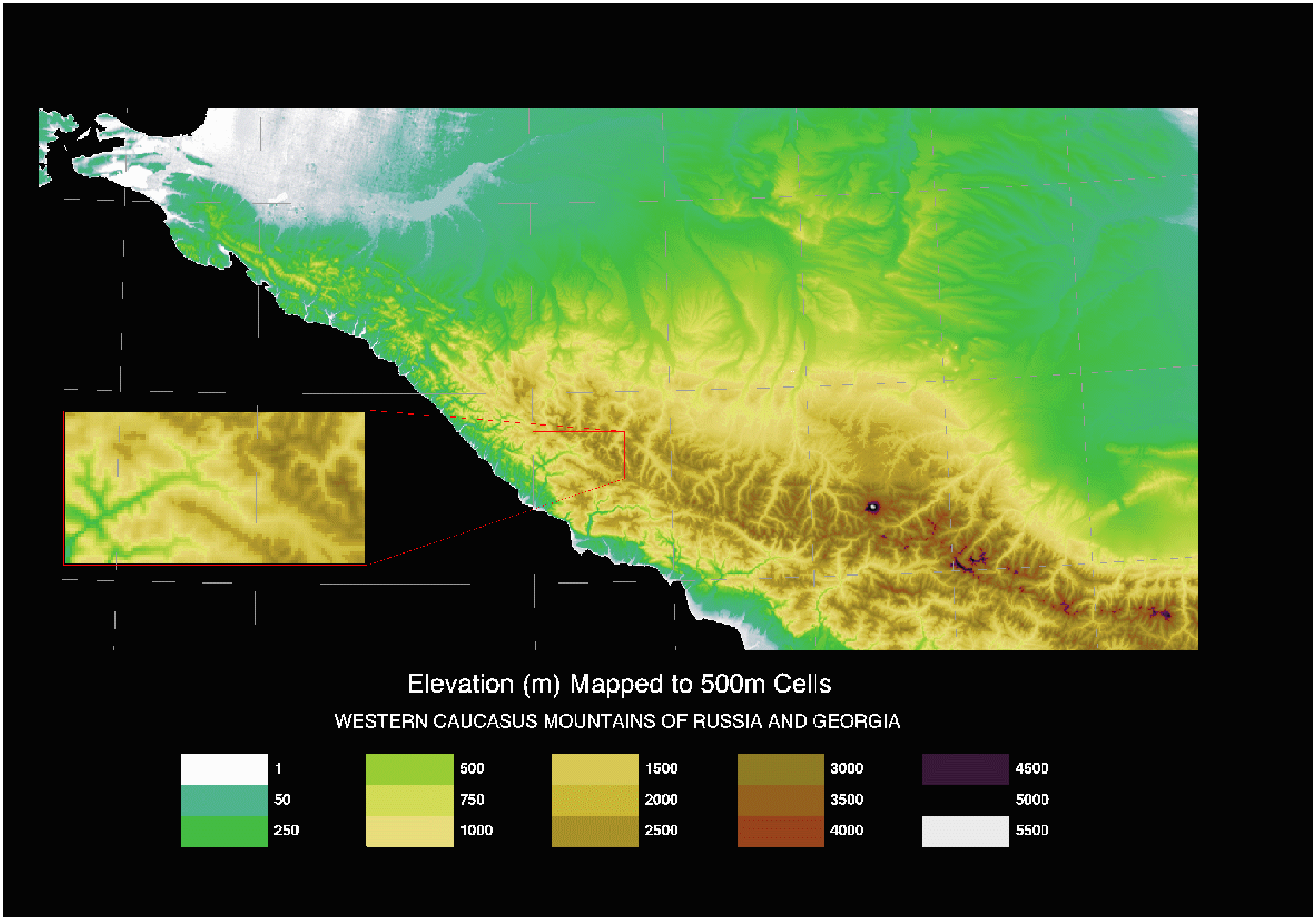

PLATTE RIVER DRAINAGE
POTENTIAL NATURAL VEGETATION
CURRENT LAND COVER
0.5 degree
0.5 degree
5 minute
5 minute
Tundra
Forests
XeroWoodlands
Savannahs
Grasslands
Shrublands
Excluded
Urban
Croplands
Cropland/Forest
Cropland/Grassland

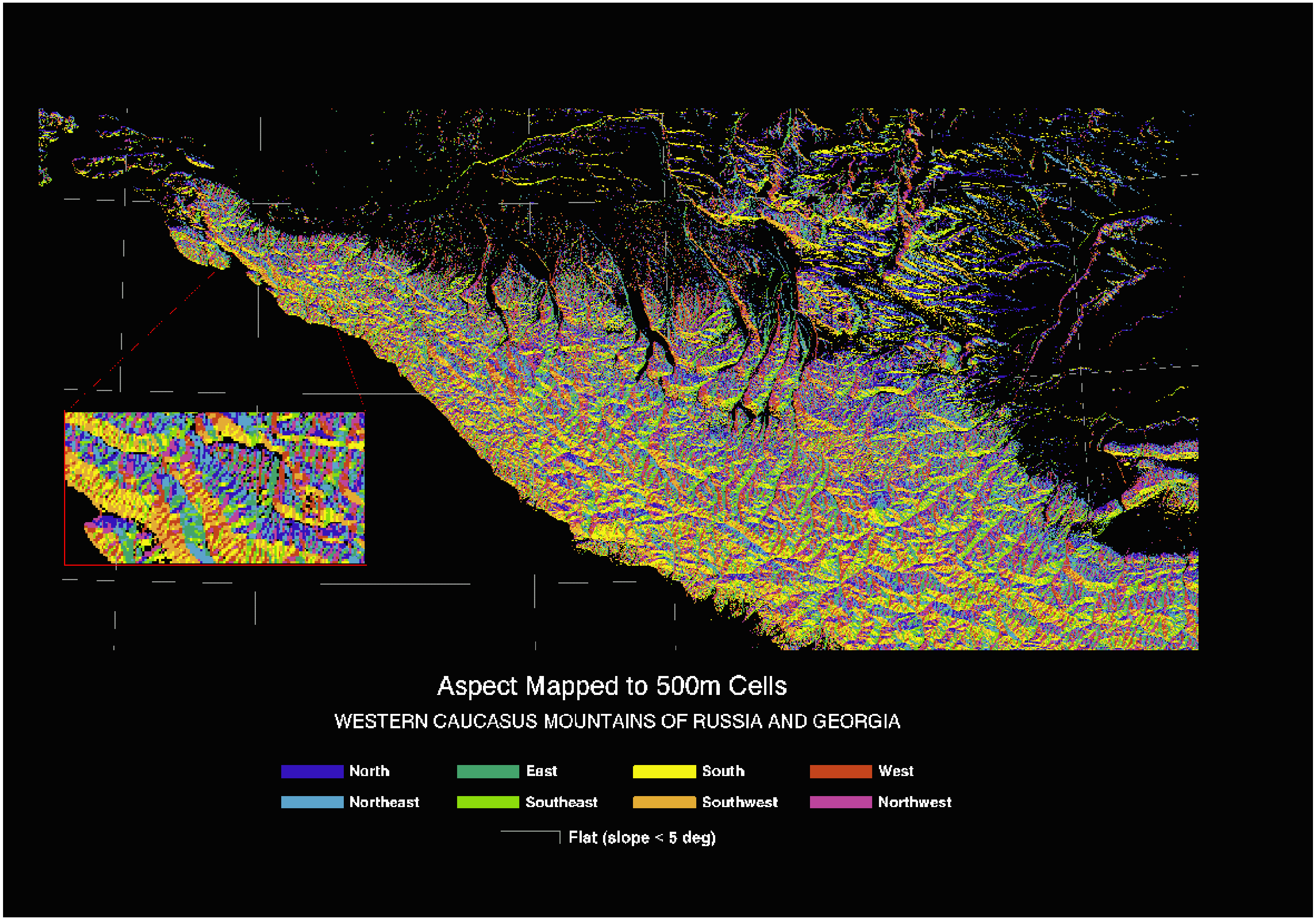

97.5W -> 98W : 39.5N -> 40N
COUNT
SOIL MODE
DOMINANT VEG CLASSES
South WW
N Central WW
S Central WW
50% C4
MidWest DC

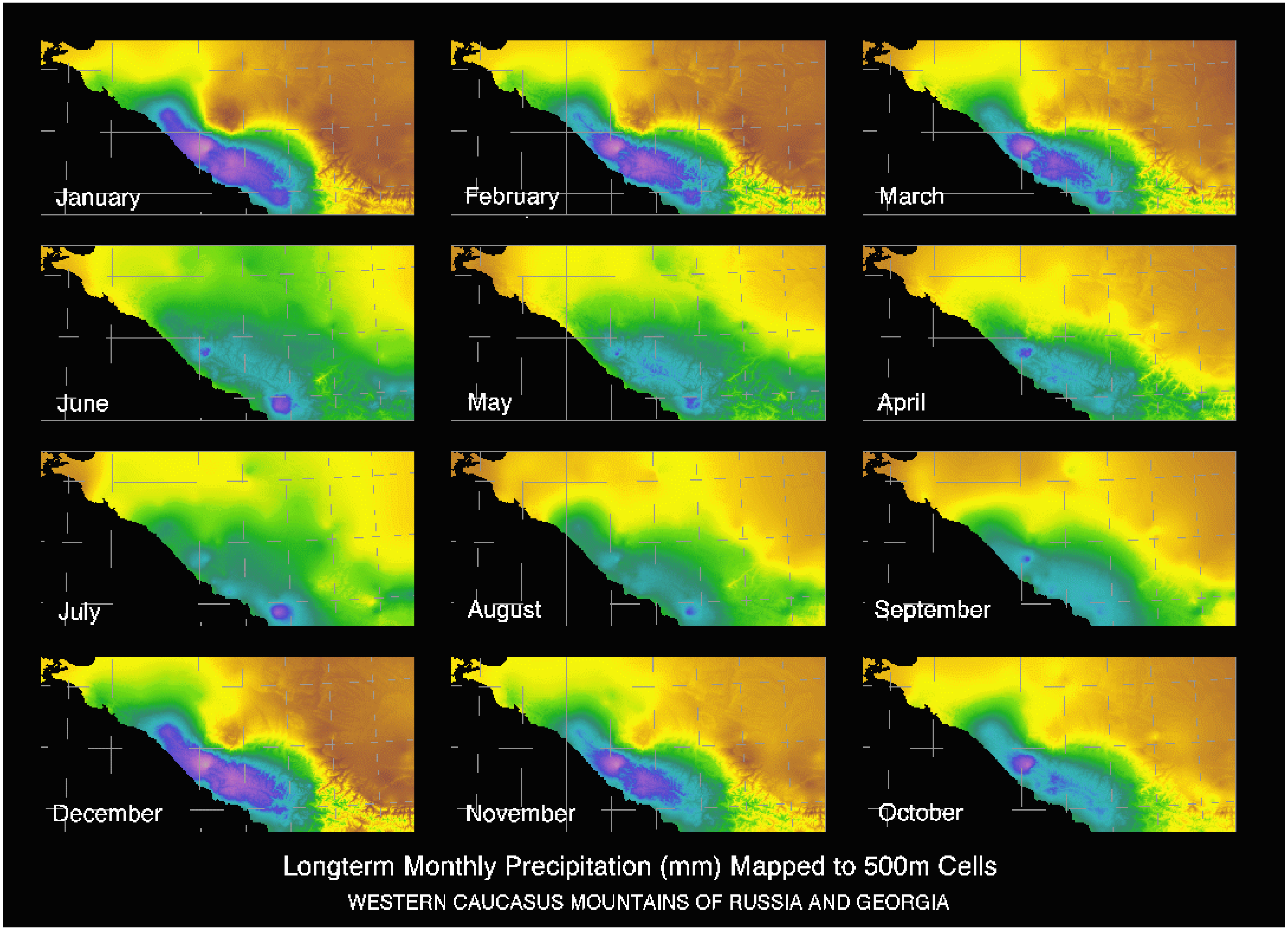

Soil Carbon
Dominant Class Potential Vegetation
Dominant Class Current Land Cover
Aggregated Subclass Current Land Cover

Net Primary Production

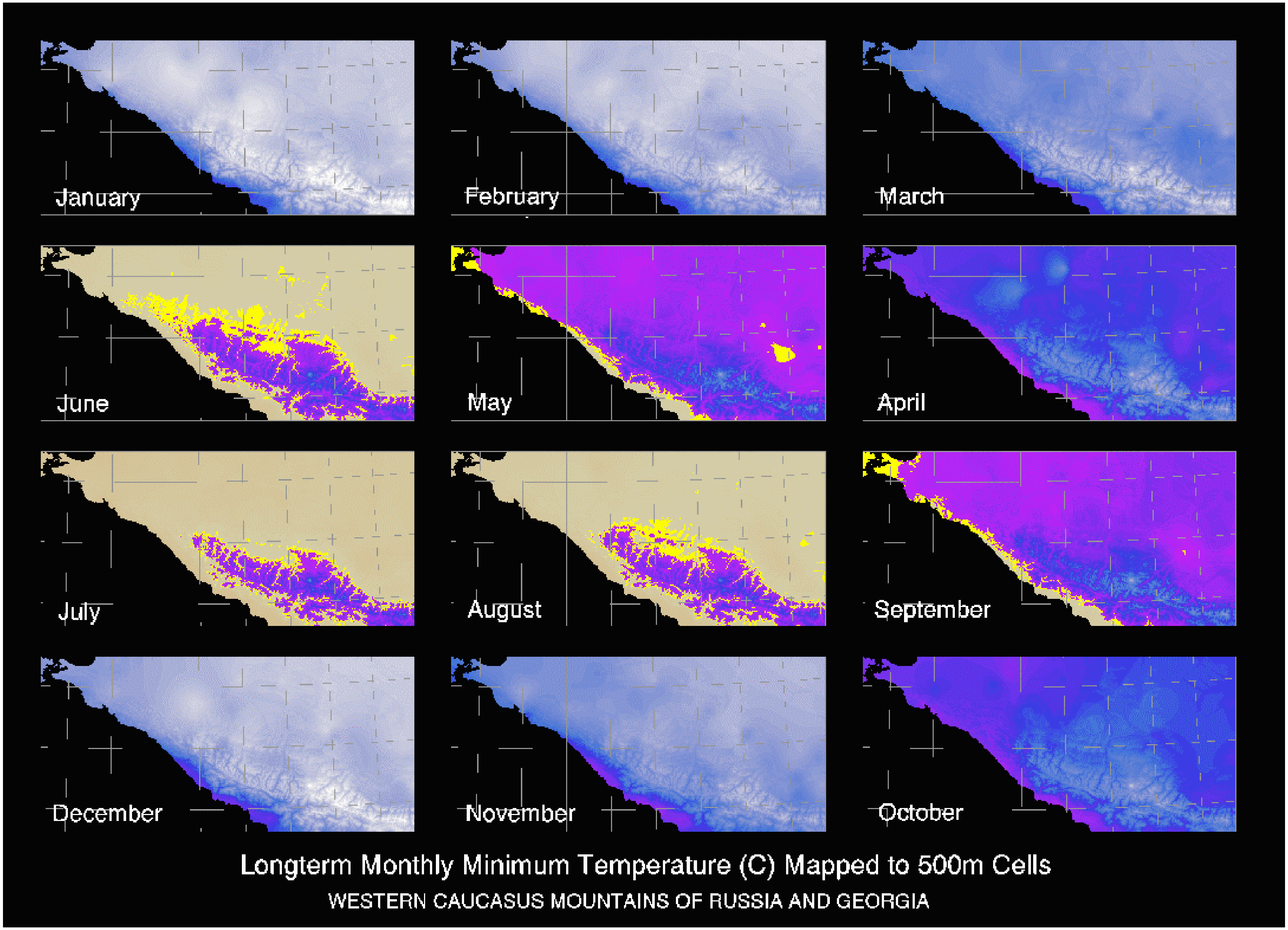



Elevation (m) Mapped to 500m Cells
WESTERN CAUCASUS MOUNTAINS OF RUSSIA AND GEORGIA
1
50
250
500
750
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500

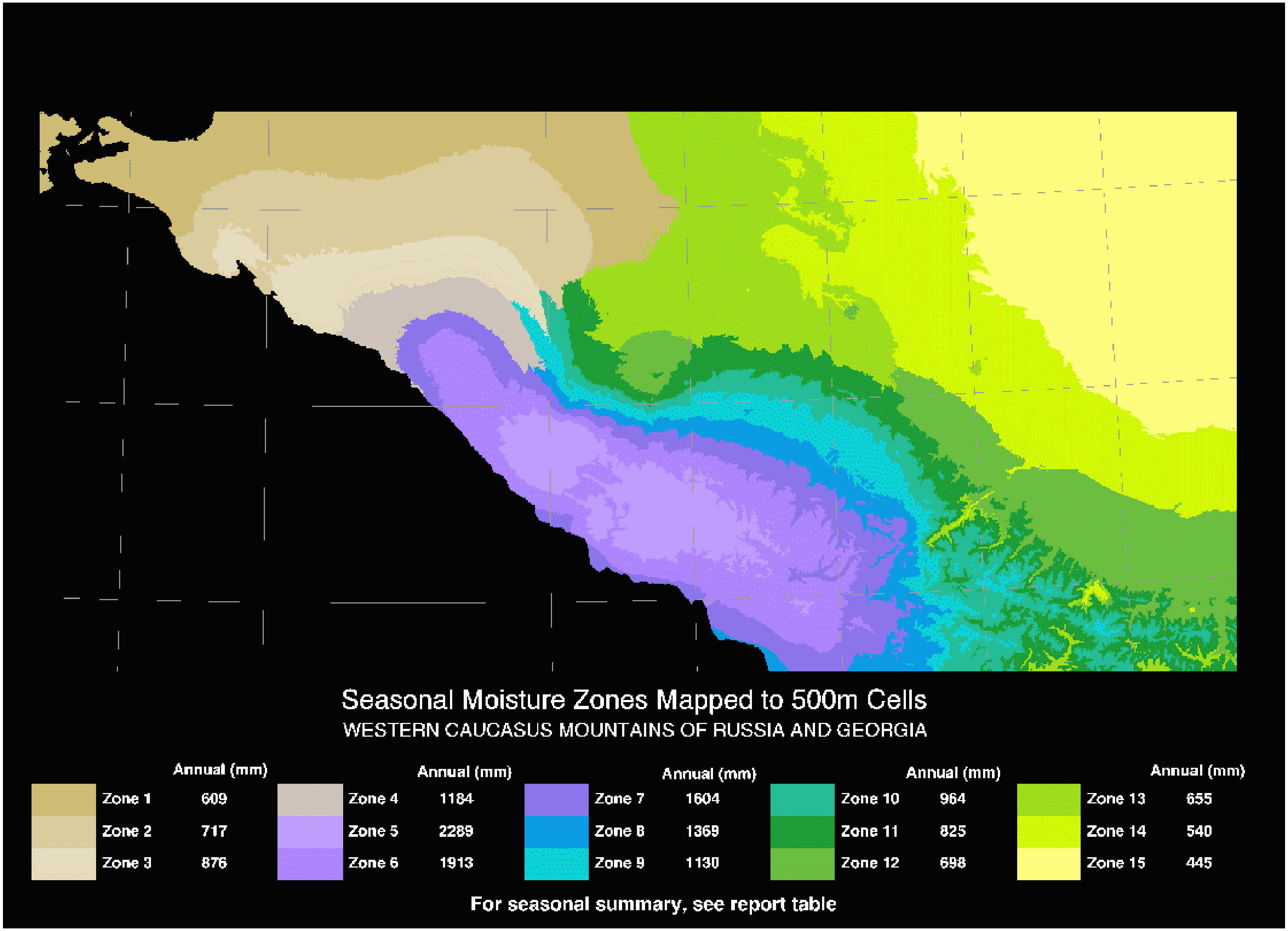

Aspect Mapped to 500m Cells
WESTERN CAUCASUS MOUNTAINS OF RUSSIA AND GEORGIA
North
East
South
West
Northeast
Southeast
Southwest
Northwest
Flat (slope < 5 deg)

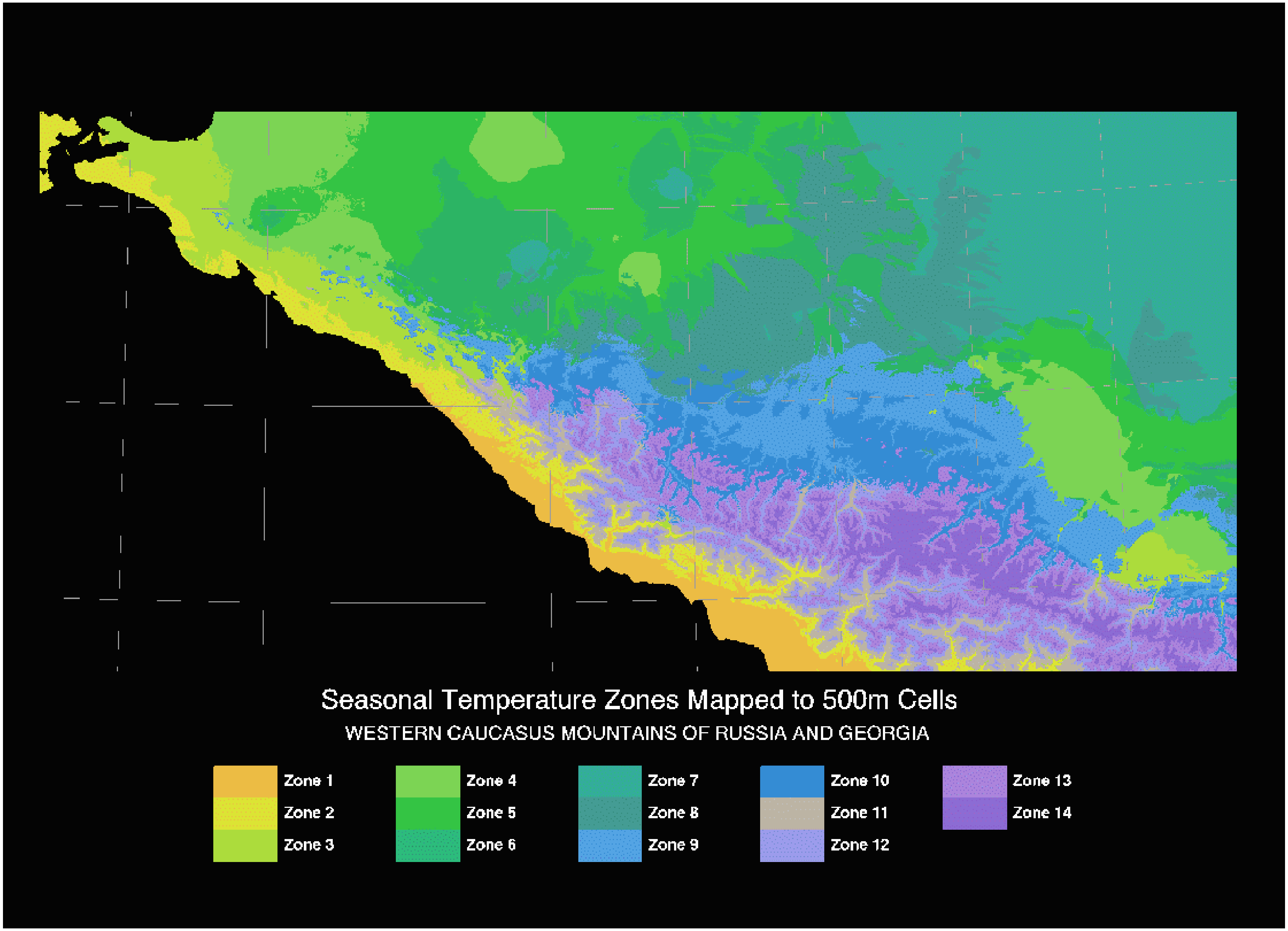

January
February
March
June
May
April
July
August
September
December
November
October
Longterm Monthly Precipitation (mm) Mapped to 500m Cells
WESTERN CAUCASUS MOUNTAINS OF RUSSIA AND GEORGIA

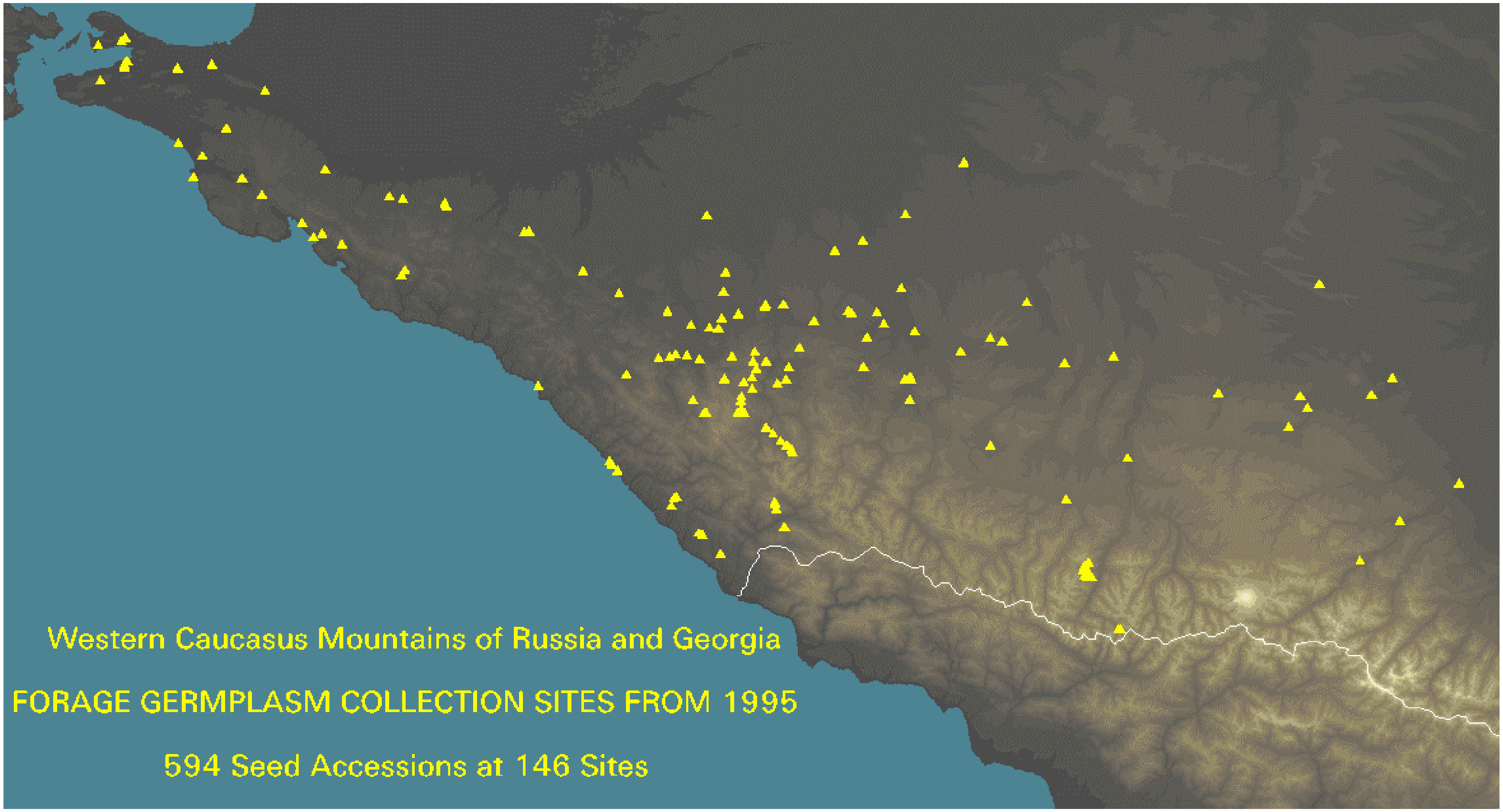

25
50
75
100
125
150
175
200
225
250
300
350
400
450


January
February
March
June
May
April
July
August
September
December
November
October
Longterm Monthly Minimum Temperature (C) Mapped to 500m Cells
WESTERN CAUCASUS MOUNTAINS OF RUSSIA AND GEORGIA


-50
-40
-30
-20
-10
-5
0
5
10


Seasonal Moisture Zones Mapped to 500m Cells
WESTERN CAUCASUS MOUNTAINS OF RUSSIA AND GEORGIA
Annual (mm)
Zone 1 609
Zone 2 717
Zone 3 876
Zone 4 1184
Zone 5 2289
Zone 6 1913
Zone 7 1604
Zone 8 1369
Zone 9 1130
Zone 10 964
Zone 11 825
Zone 12 698
Zone 13 655
Zone 14 540
Zone 15 445
For seasonal summary, see report table


Seasonal Temperature Zones Mapped to 500m Cells
WESTERN CAUCASUS MOUNTAINS OF RUSSIA AND GEORGIA
Zone 1
Zone 2
Zone 3
Zone 4
Zone 5
Zone 6
Zone 7
Zone 8
Zone 9
Zone 10
Zone 11
Zone 12
Zone 13
Zone 14


Western Caucasus Mountains of Russia and Georgia
FORAGE GERMPLASM COLLECTION SITES FROM 1995
594 Seed Accessions at 146 Sites

PCRaster (1995) contact pcraster@frw.ruu.nl how to get version 1 of PCRaster. Version 2, used in this paper, will be available by spring 1996.

Fig 1 Map of the research area

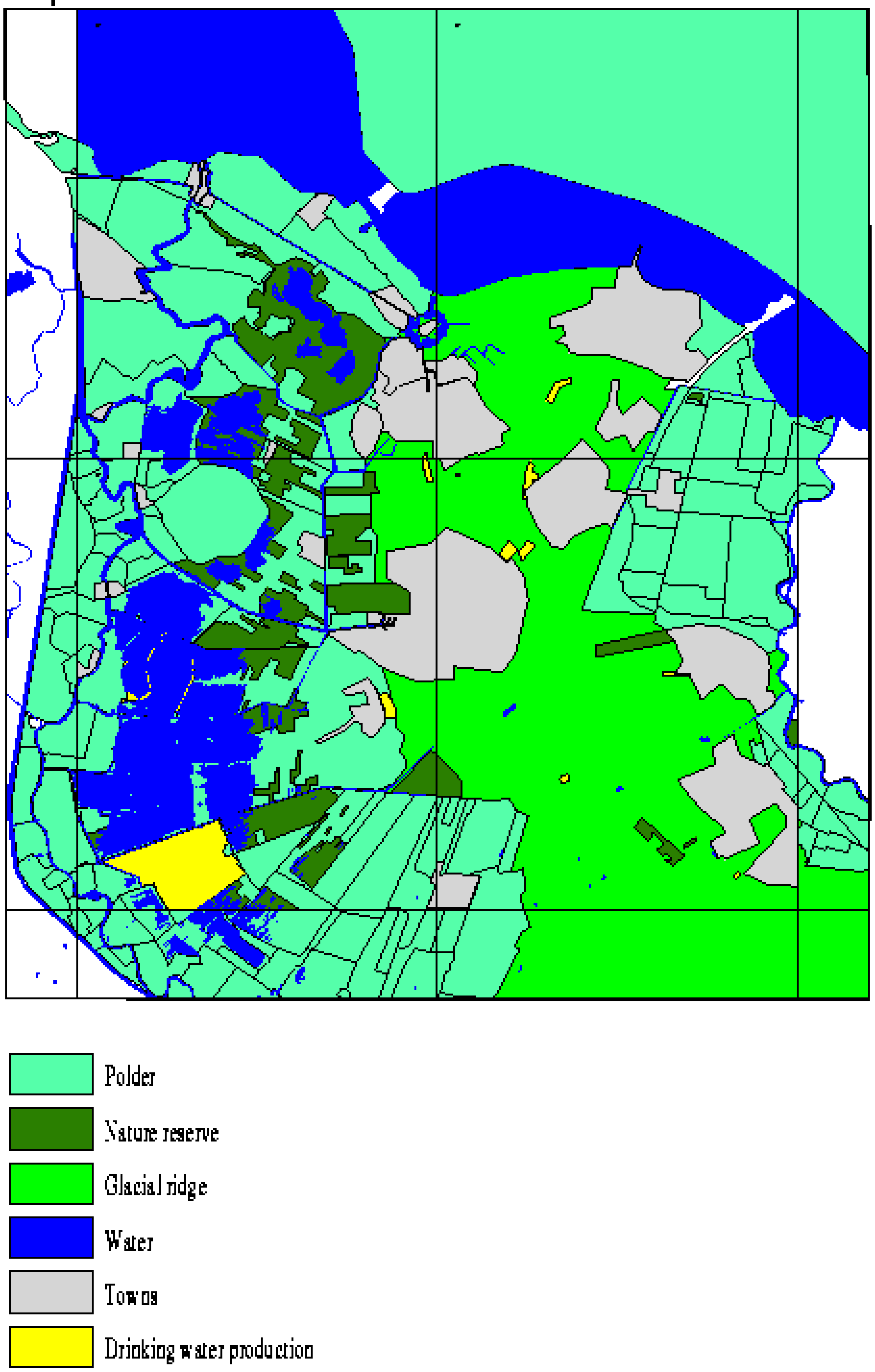


Fig 2. Map with data points of chloride in groundwater

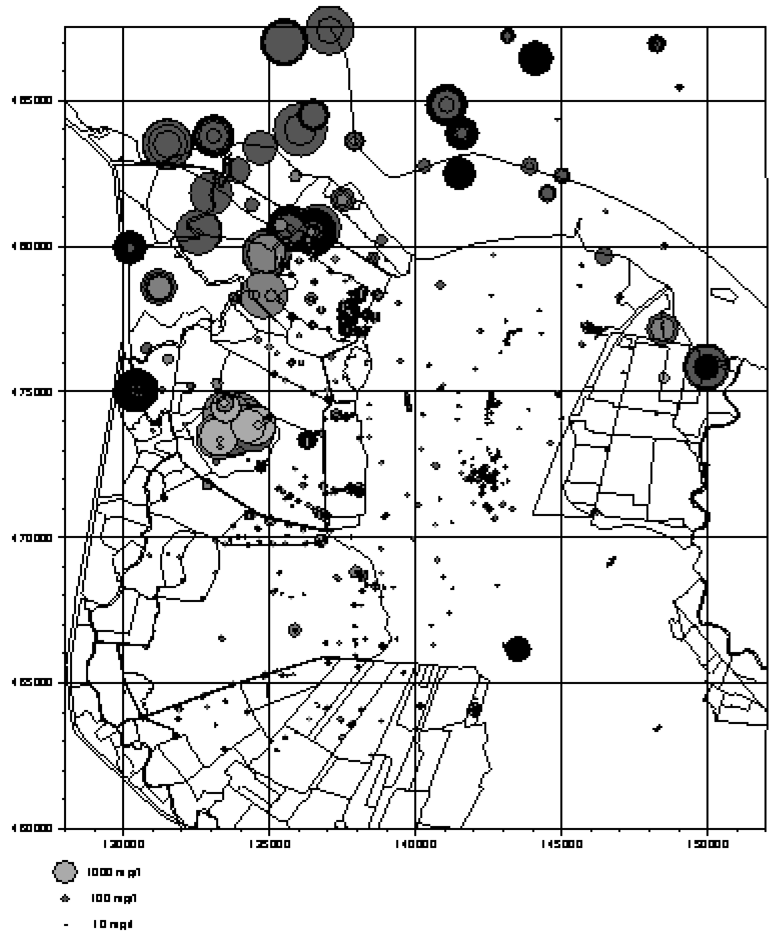


Fig 3. Map with data points of ammonium in groundwater

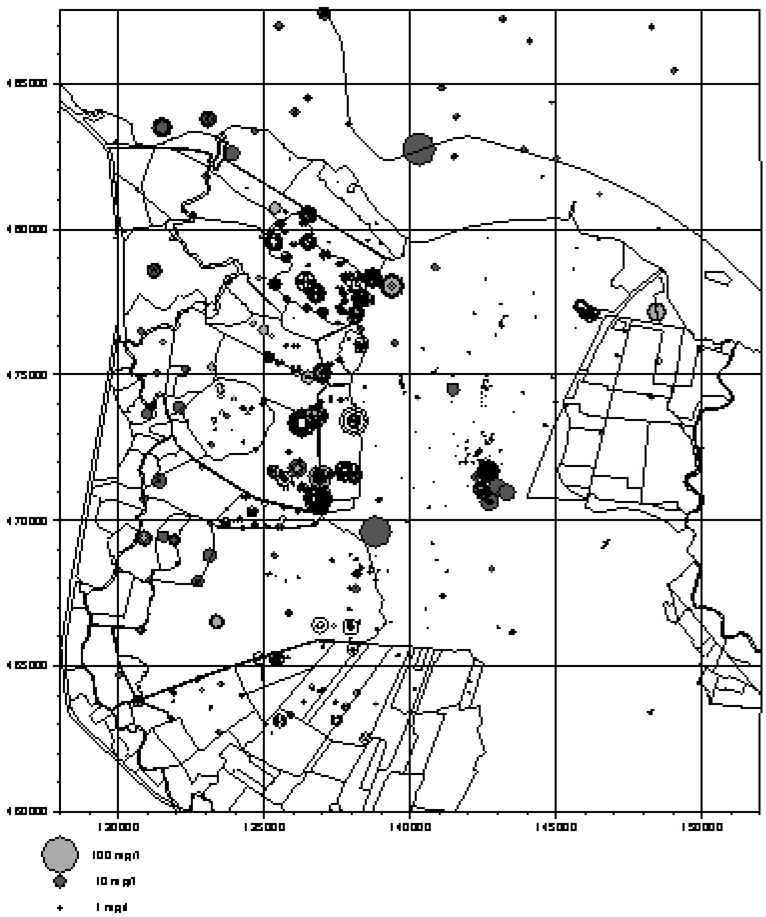


Fig 4. Estimated concentration chloride in groundwater

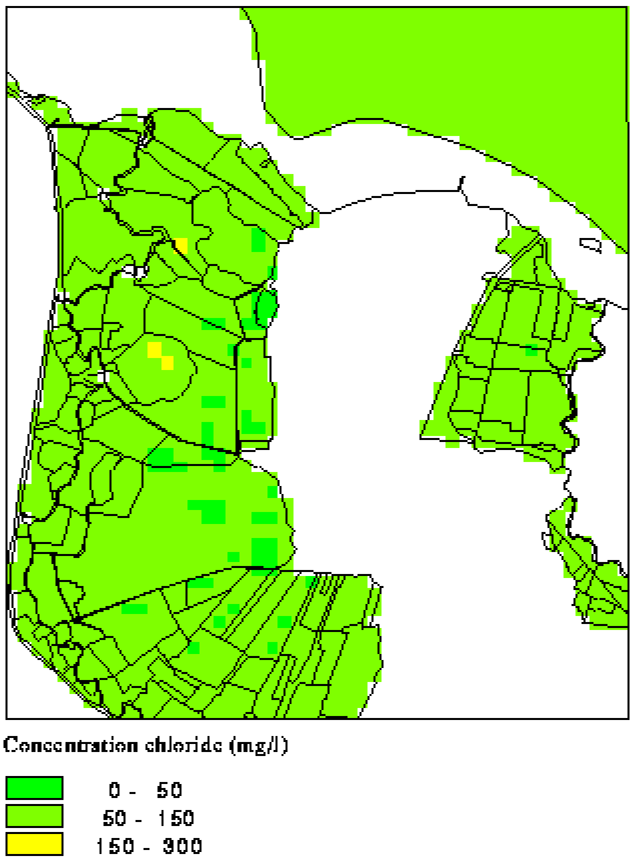


Fig 5. Estimated concentration Ammonium in groundwater

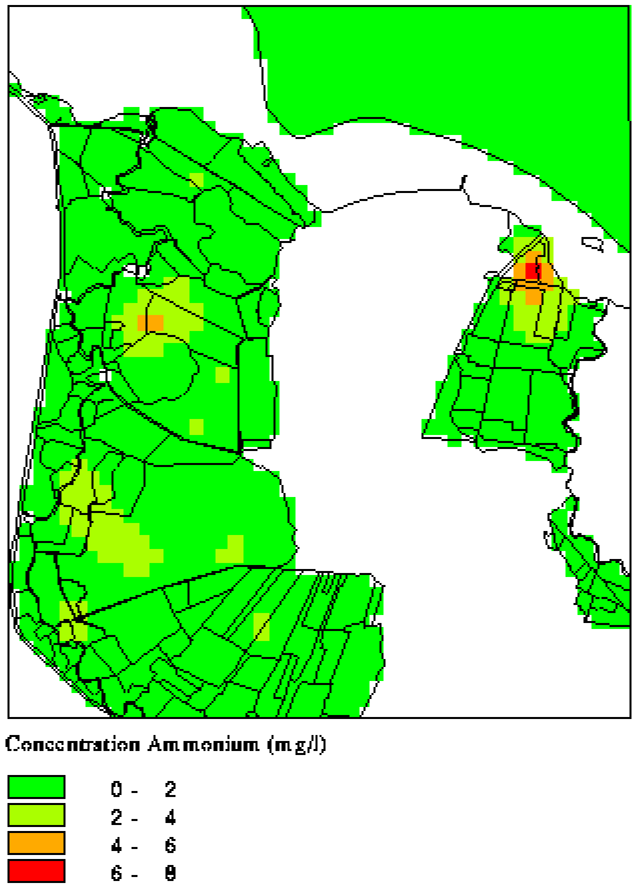


Fig 6. Maps with response values of four plant species

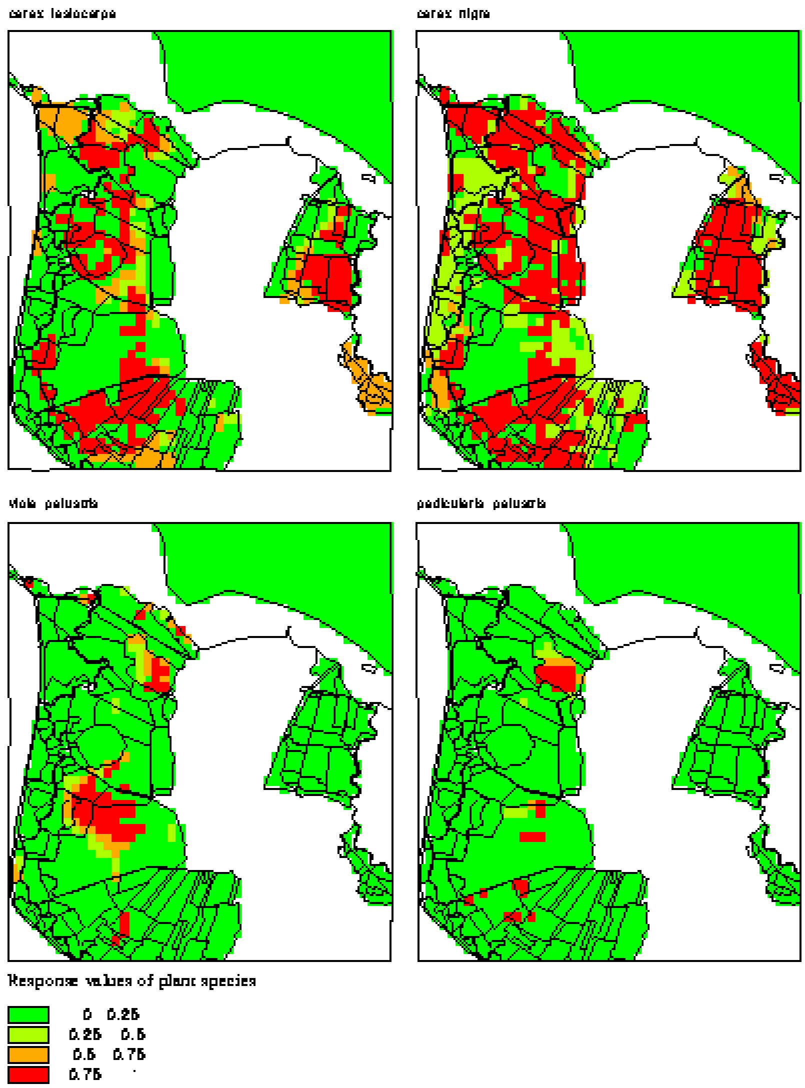


Fig 7

Map with water types

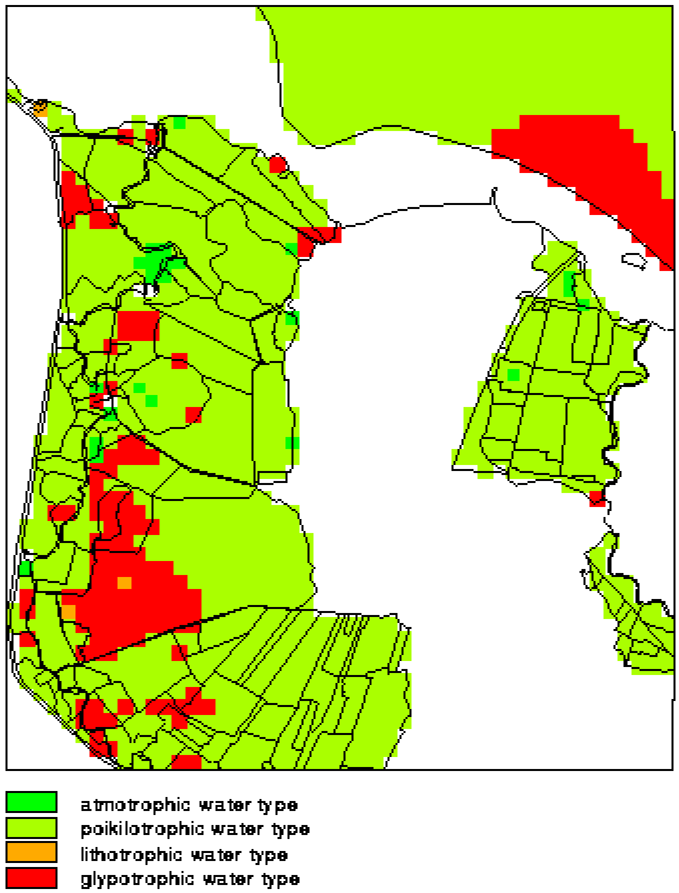


Fig 8.
Map with nutrient types

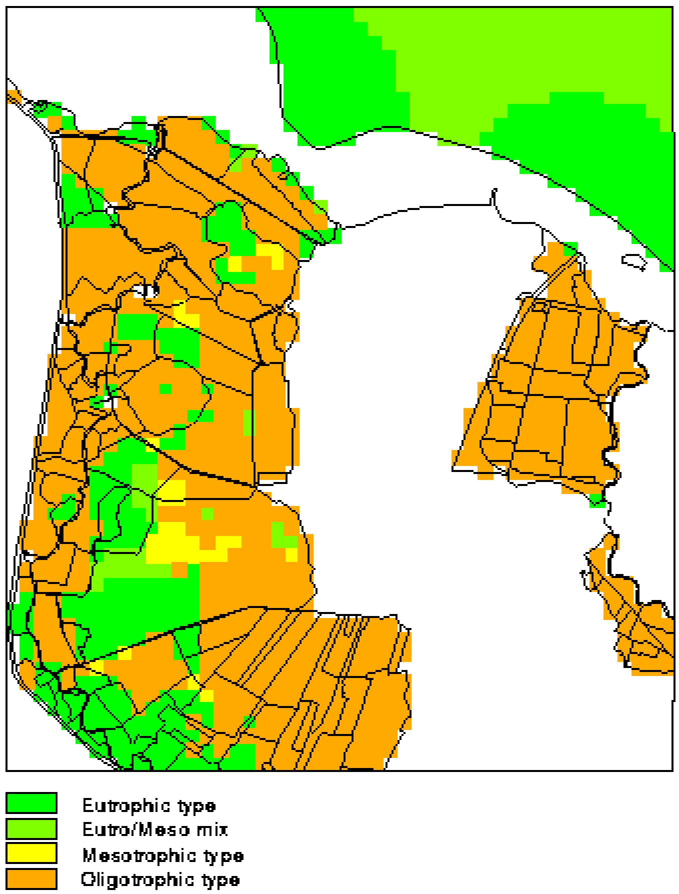


Fig 9.
Map with species groups

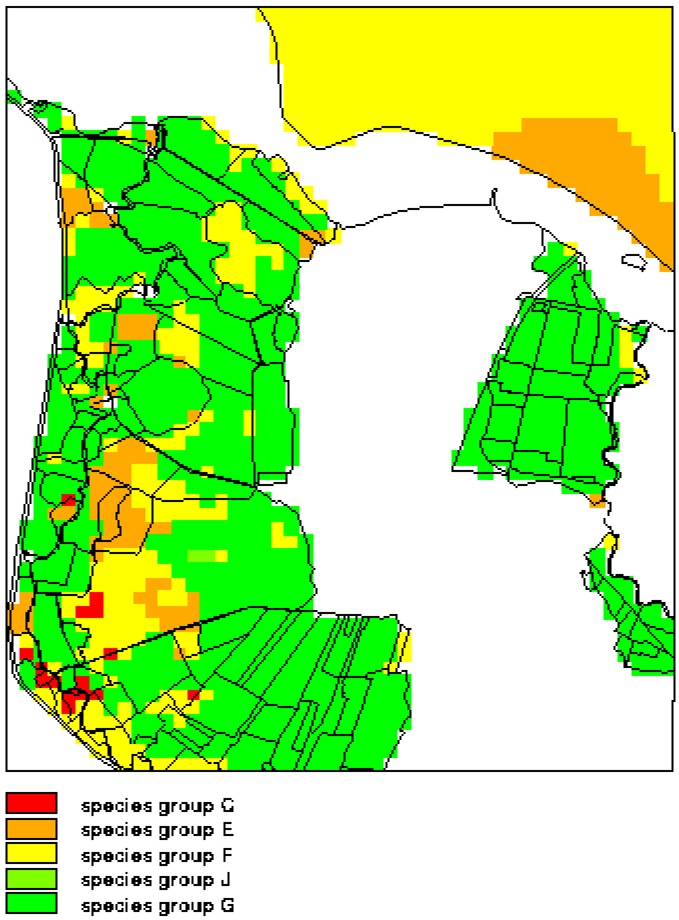


Fig 10.

All valid estimates

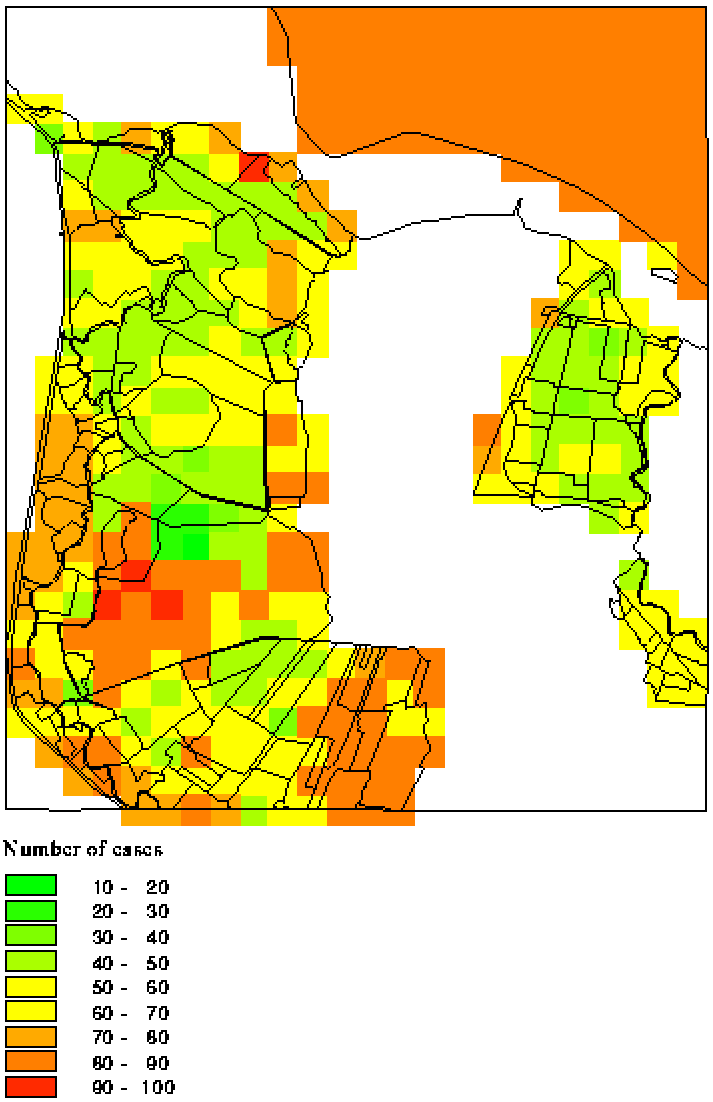
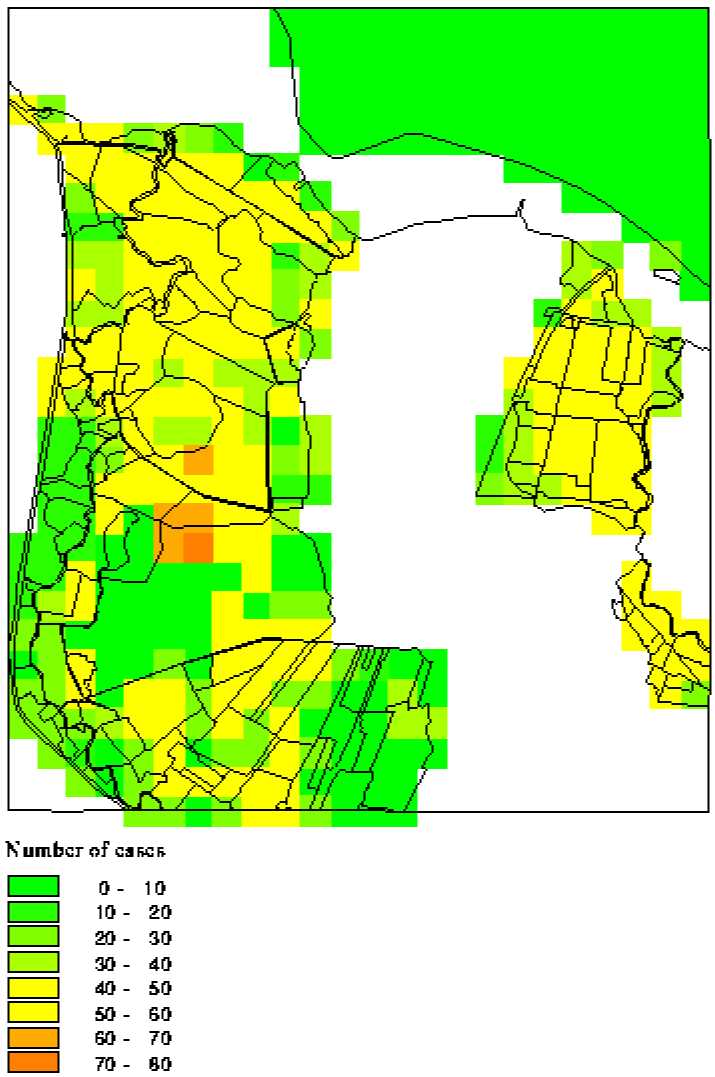
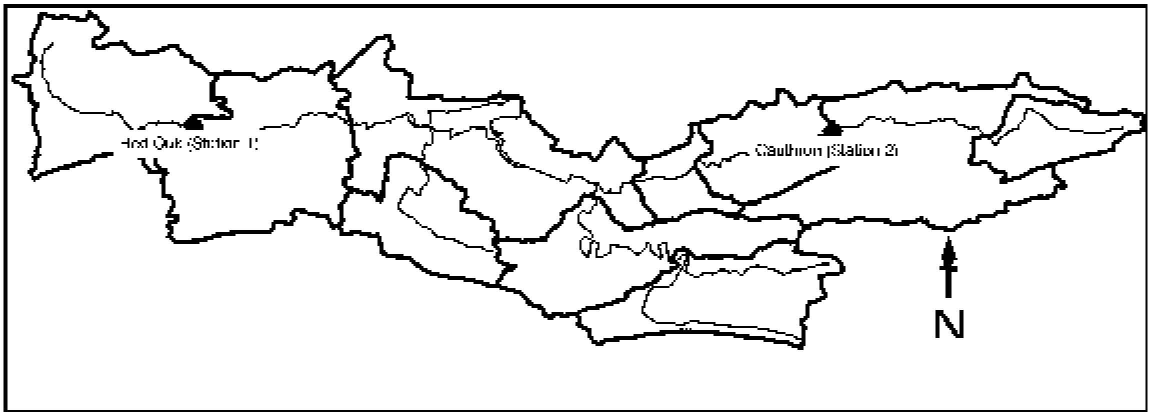
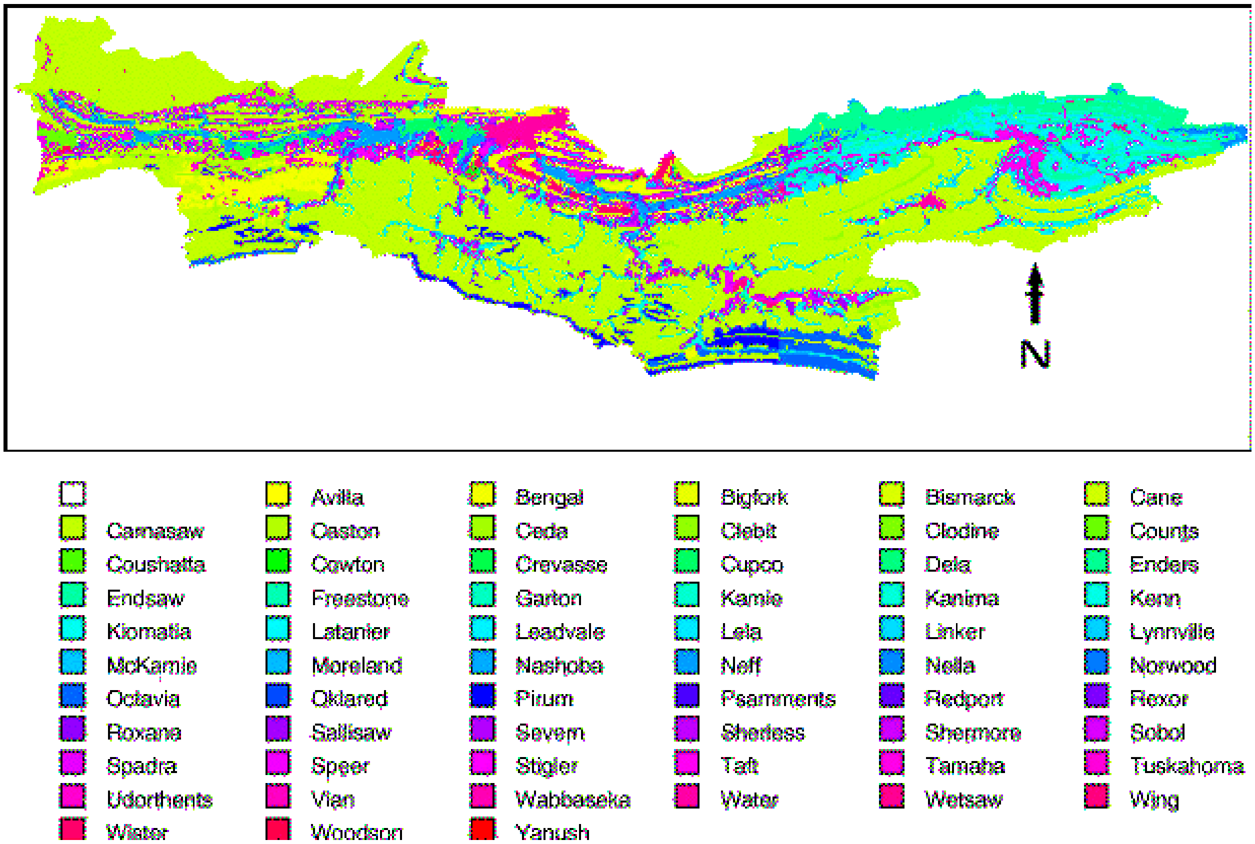


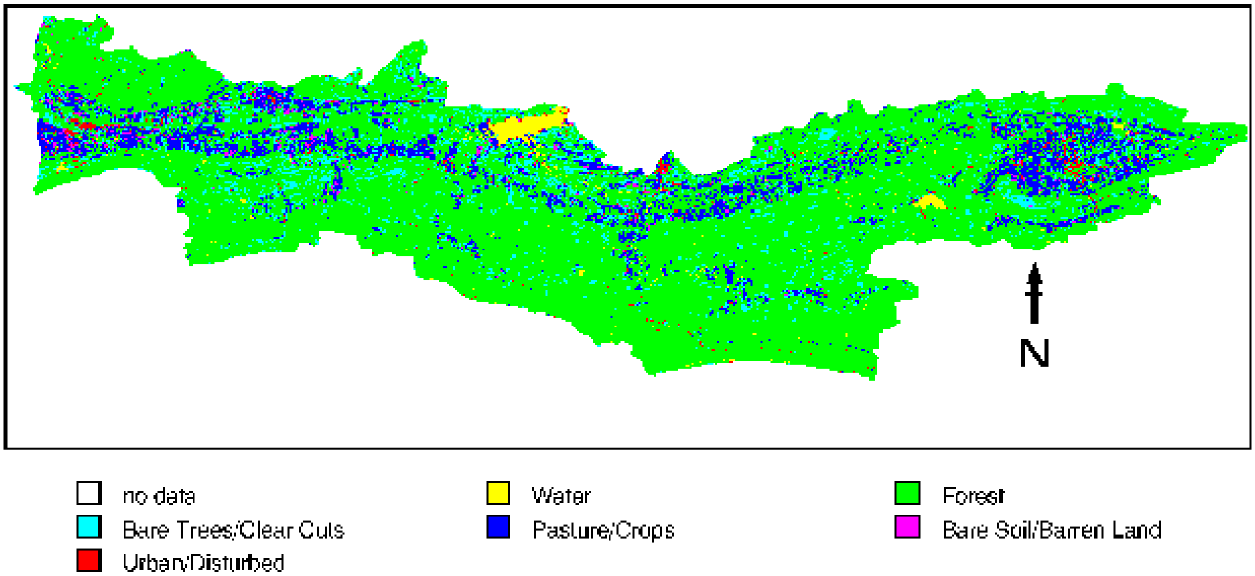
Fig 11

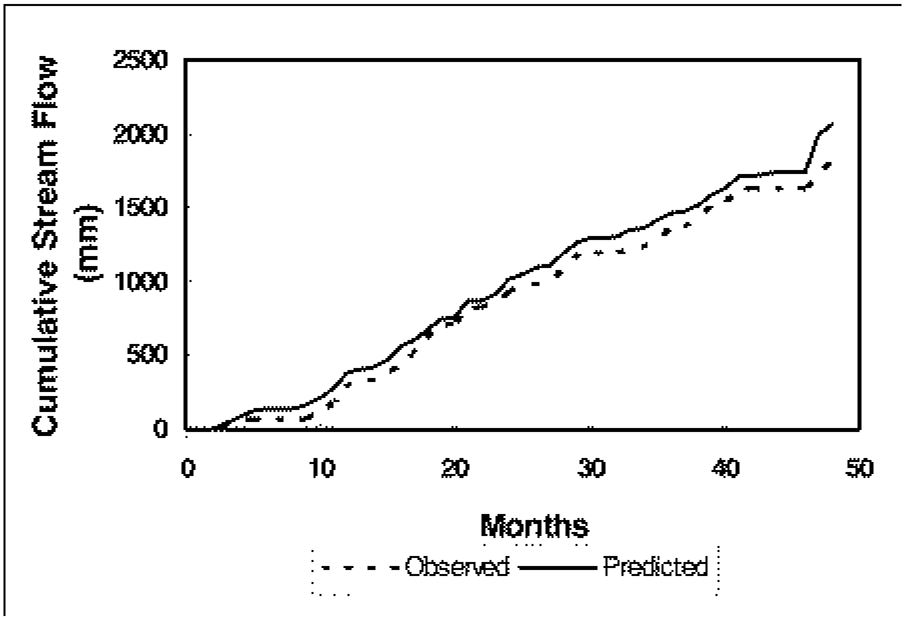
All non valid estimates

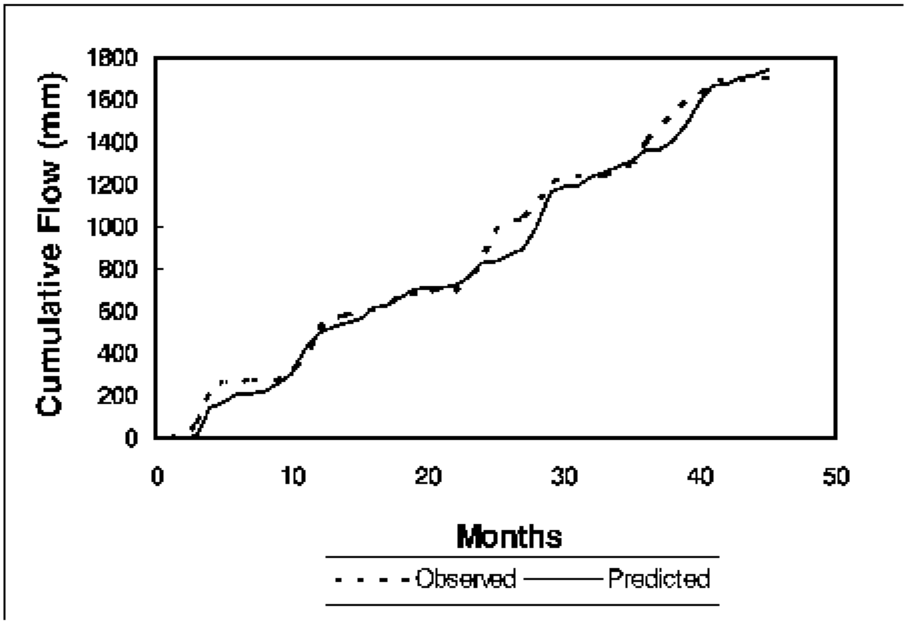


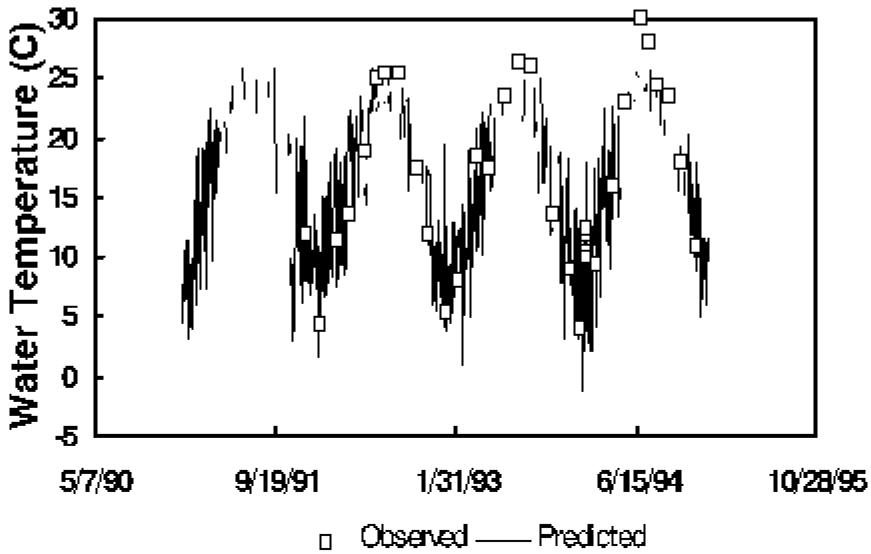

Cauthron (Station 2)
N

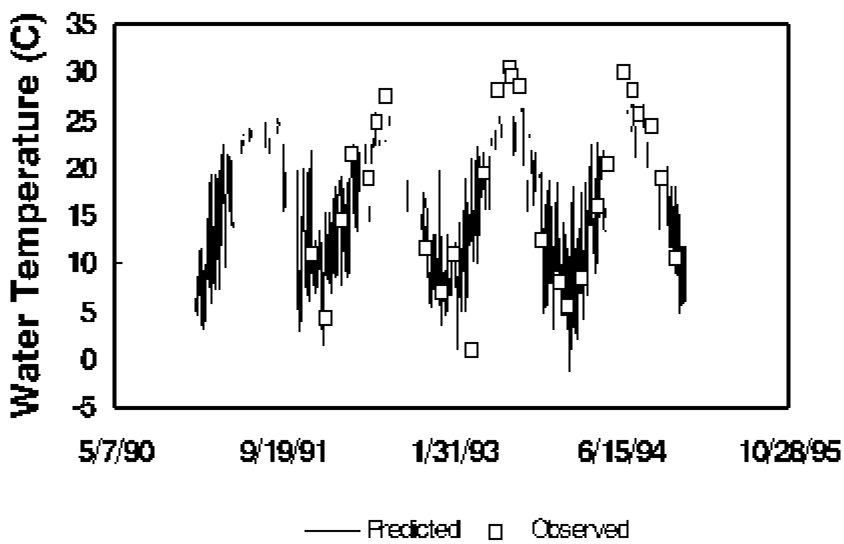

N
Avilla
Bengal
Bigfork
Bismarck
Cane
Carnasaw
Caston
Ceda
Clebit
Clodine
Counts
Coushatta
Cowton
Crevasse
Cupco
Dela
Enders
Endsaw
Freestone
Garton
Kamie
Kanima
Kenn
Kiomatia
Latanier
Leadvale
Lela
Linker
Lynnville
McKamie
Moreland
Nashoba
Neff
Nella
Norwood
Octavia
Oklared
Pirum
Psamments
Redport
Rexor
Roxana
Sallisaw
Severn
Sherless
Shermore
Sobol
Spadra
Speer
Stigler
Taft
Tamaha
Tuskahoma
Udorthents
Vian
Wabbaseka
Water
Wetsaw
Wing
Wister
Woodson
Yanush

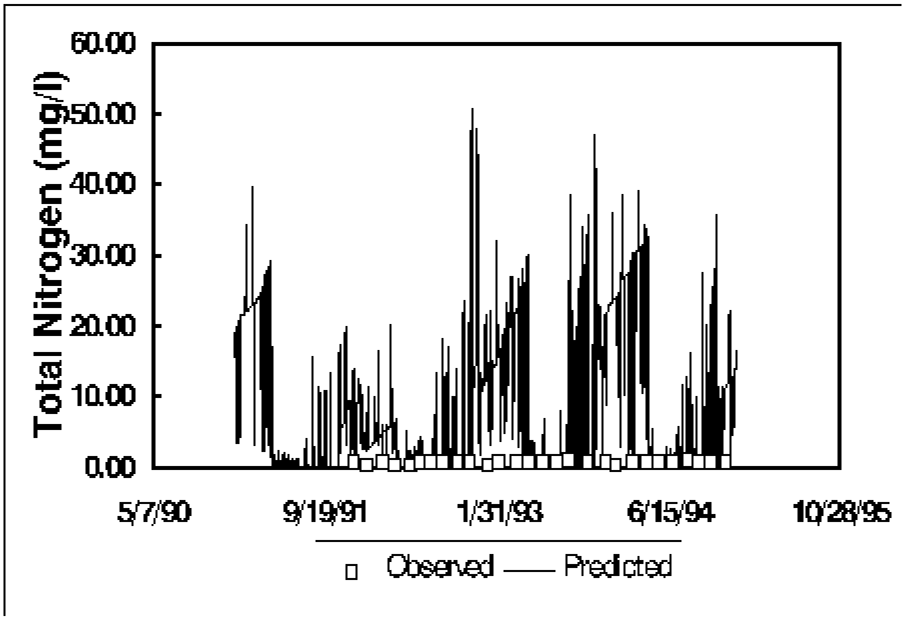

N
no data
Water
Forest
Bare Trees/Clear Cuts
Pasture/Crops
Bare Soil/Barren Land
Urban/Disturbed

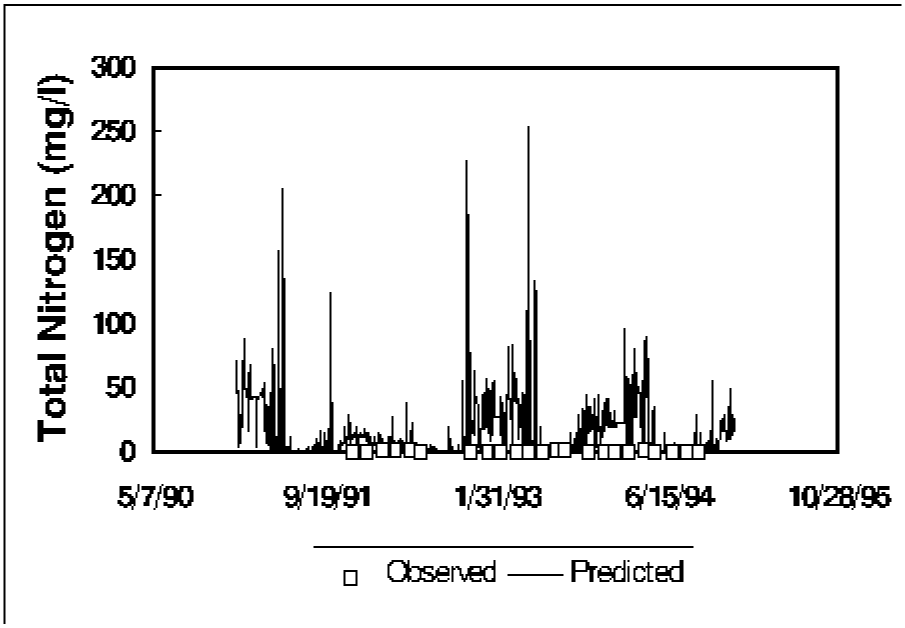

Cumulative Stream Flow (mm)
2500
2000
1500
1000
500
0
0
10
20
30
40
50
Months
Observed
Predicted

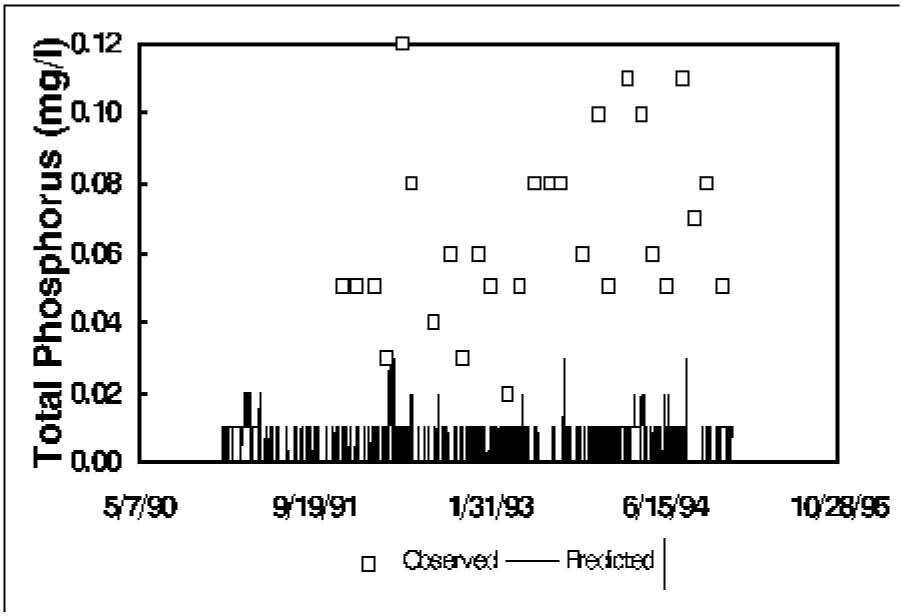

Cumulative Flow (mm)
1800
1600
1400
1200
1000
800
600
400
200
0
0
10
20
30
40
50
Months
- - - - Observed
Predicted

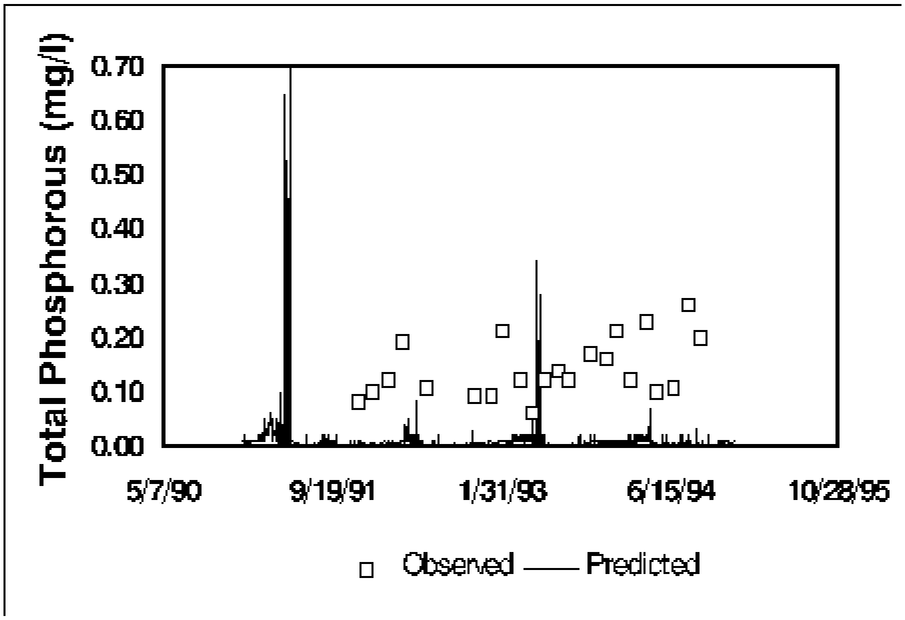

Water Temperature (C)
30
25
20
15
10
5
0
-5
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
Observed
Predicted

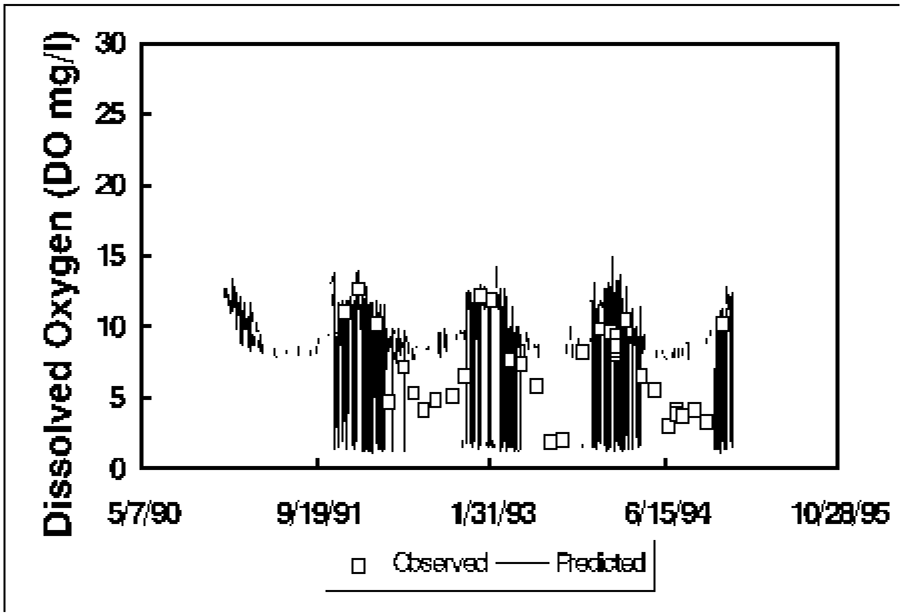

Water Temperature (C)
35
30
25
20
15
10
5
0
-5
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
Predicted
Observed

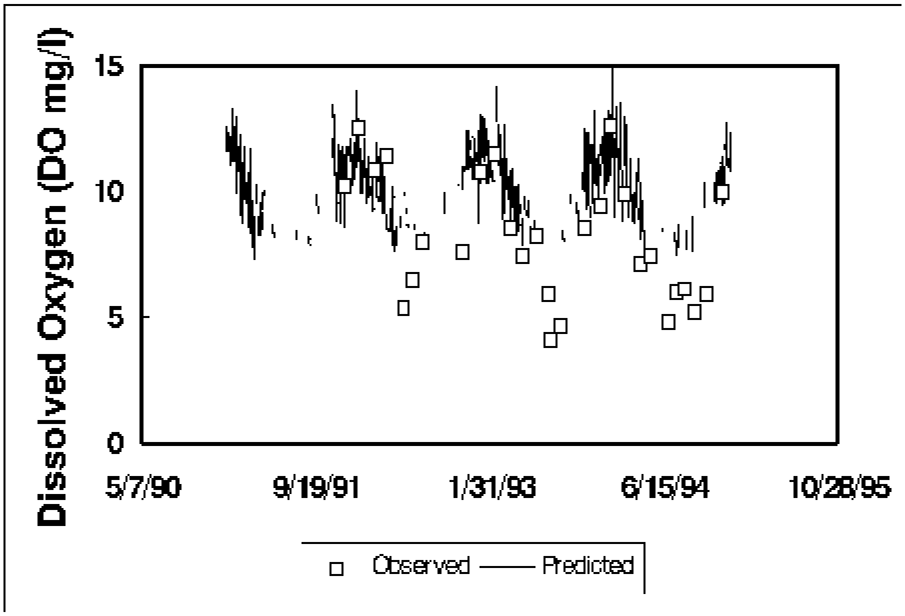

Total Nitrogen (mg/l)
60.00
50.00
40.00
30.00
20.00
10.00
0.00
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
Observed
Predicted


Total Nitrogen (mg/l)
300
250
200
150
100
50
0
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
Observed
Predicted

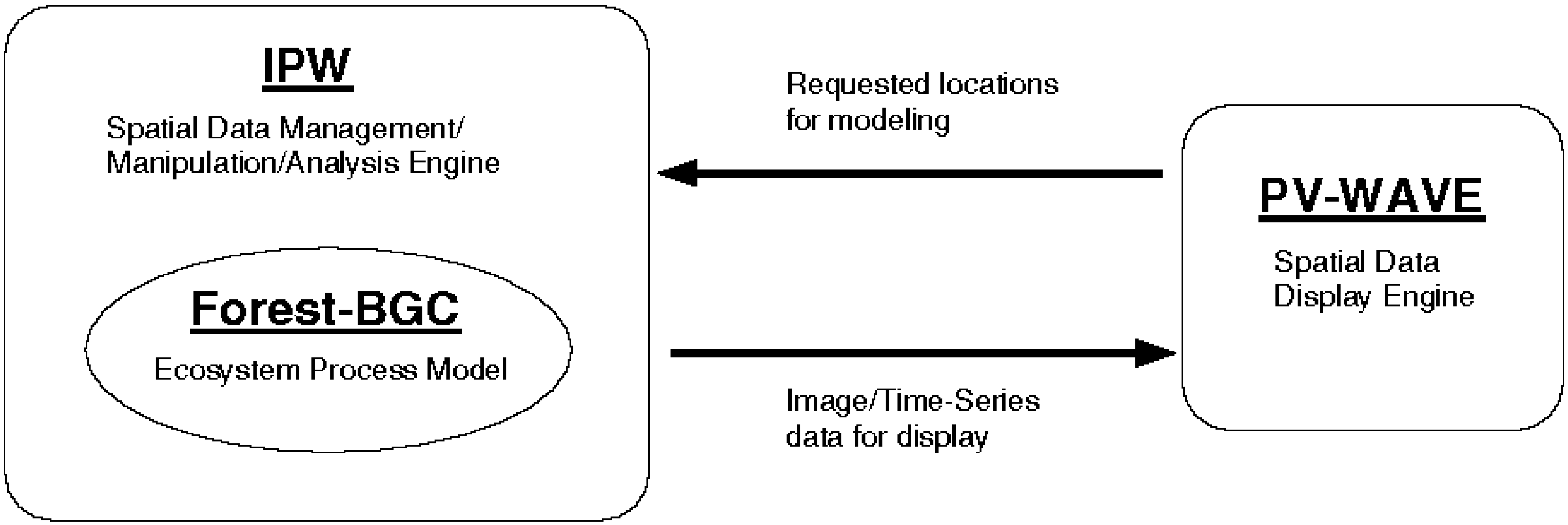

Total Phosphorus (mg/l)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
Observed
Predicted

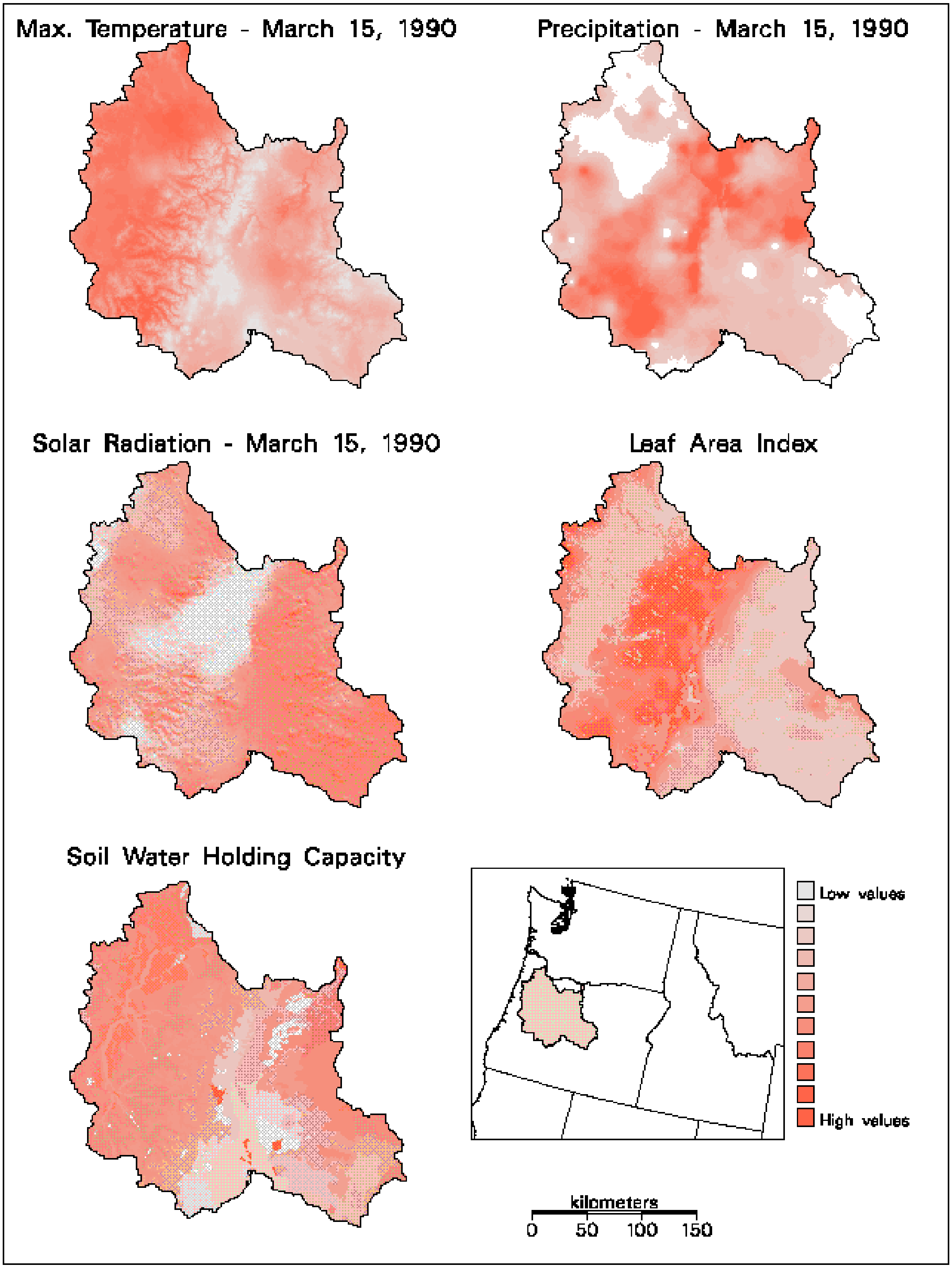

Total Phosphorous (mg/l)
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
Observed
Predicted

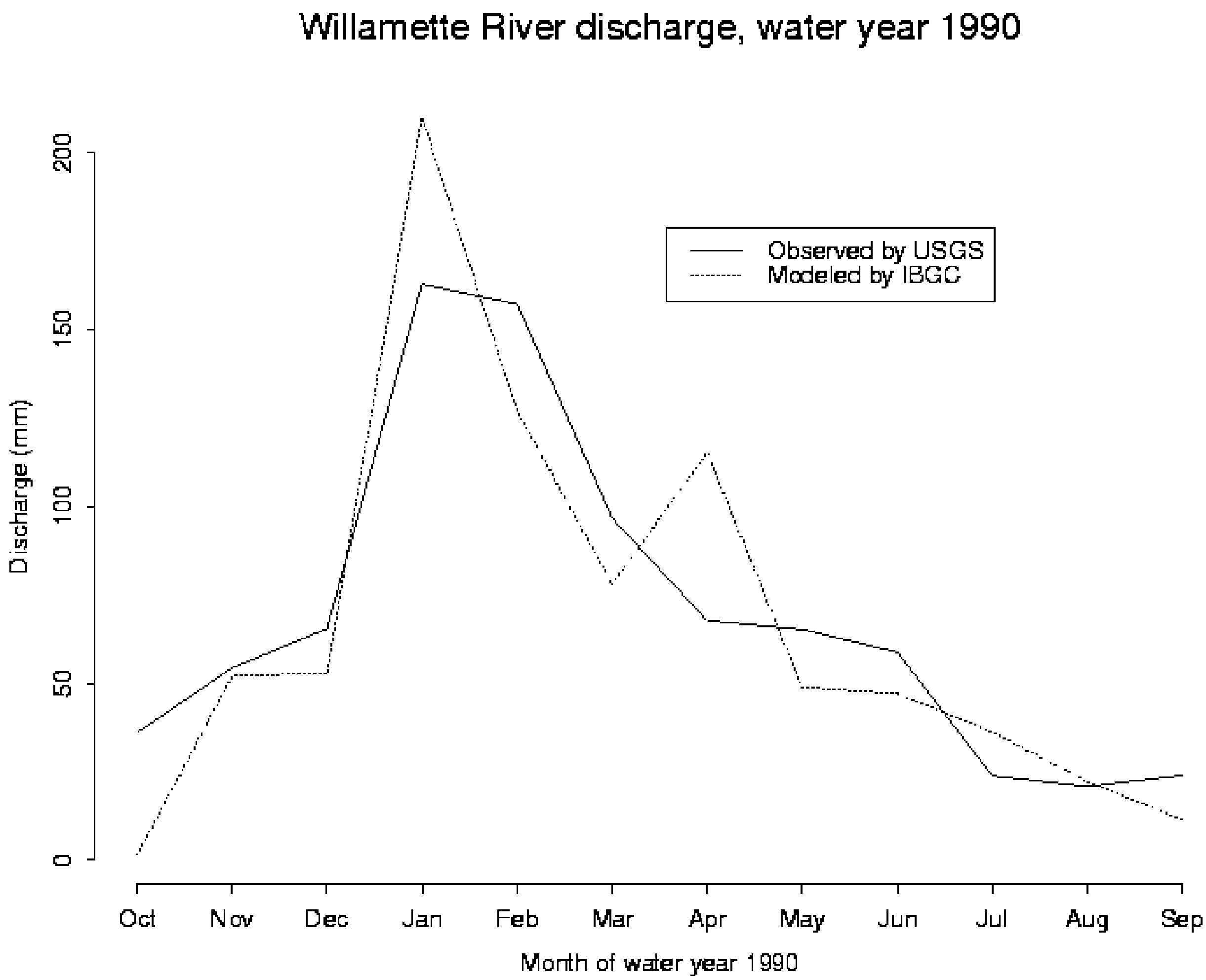

Dissolved Oxygen (DO mg/l)
30
25
20
15
10
5
0
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
Observed
Predicted

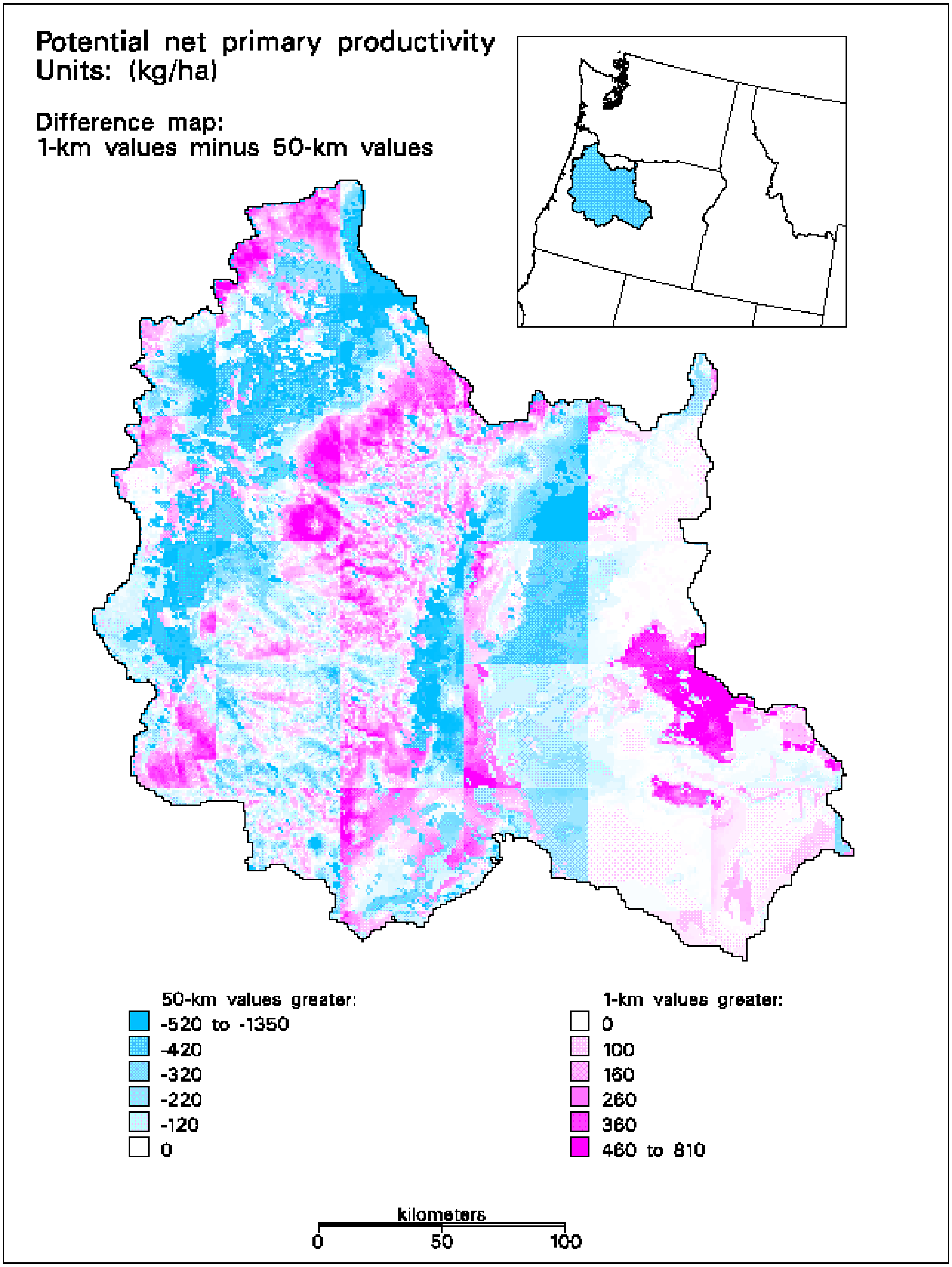

Dissolved Oxygen (DO mg/l)
15
10
5
0
5/7/90
9/19/91
1/31/93
6/15/94
10/28/95
Observed
Predicted

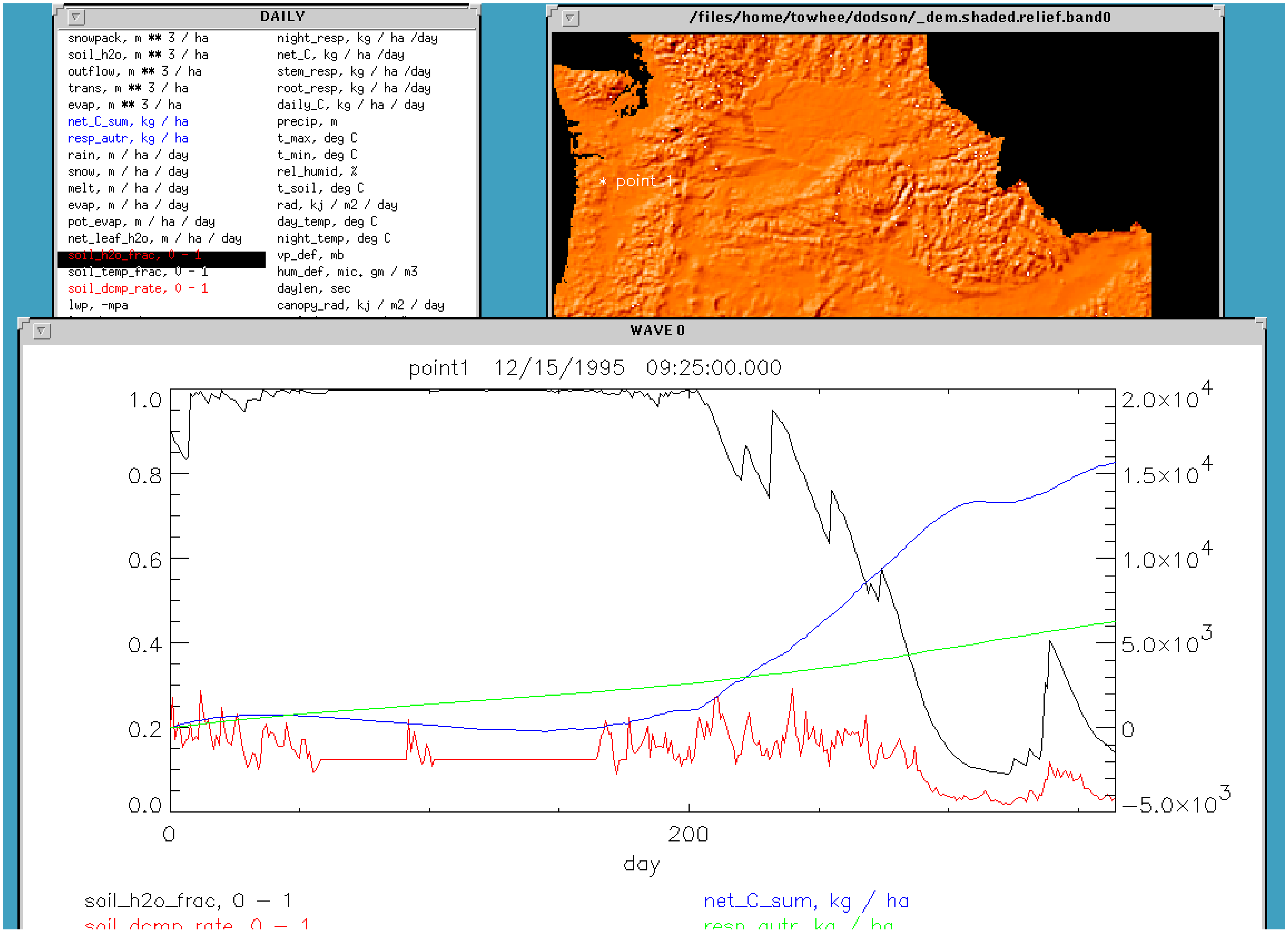
Integrated Spatial Modeling Framework

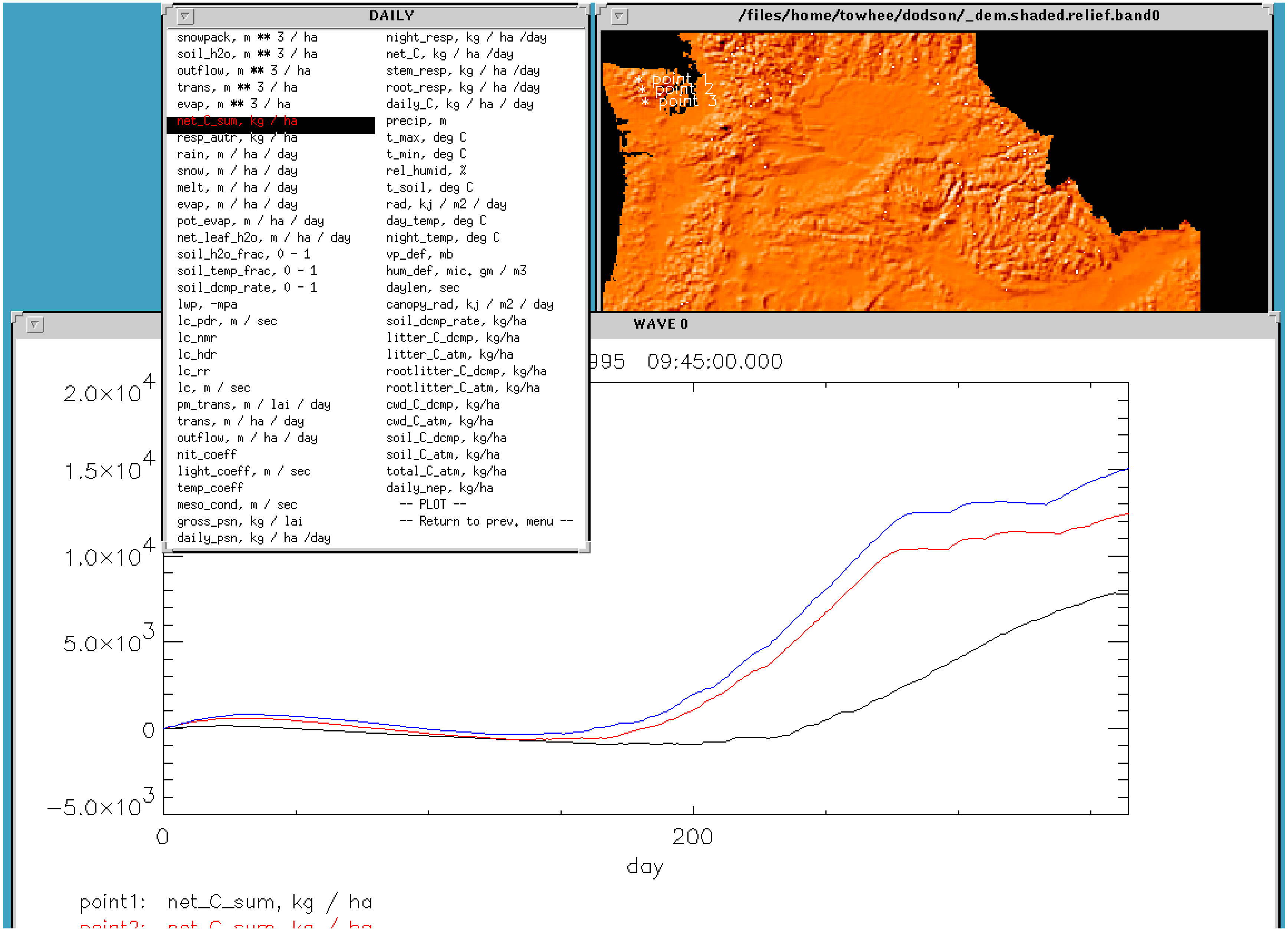


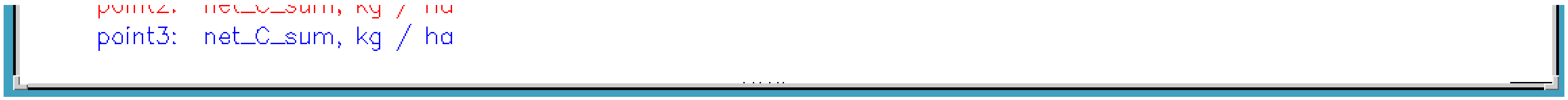

Max. Temperature - March 15, 1990
Precipitation - March 15, 1990
Solar Radiation - March 15, 1990
Leaf Area Index
Soil Water Holding Capacity
Low values
High values
kilometers
0 50 100 150

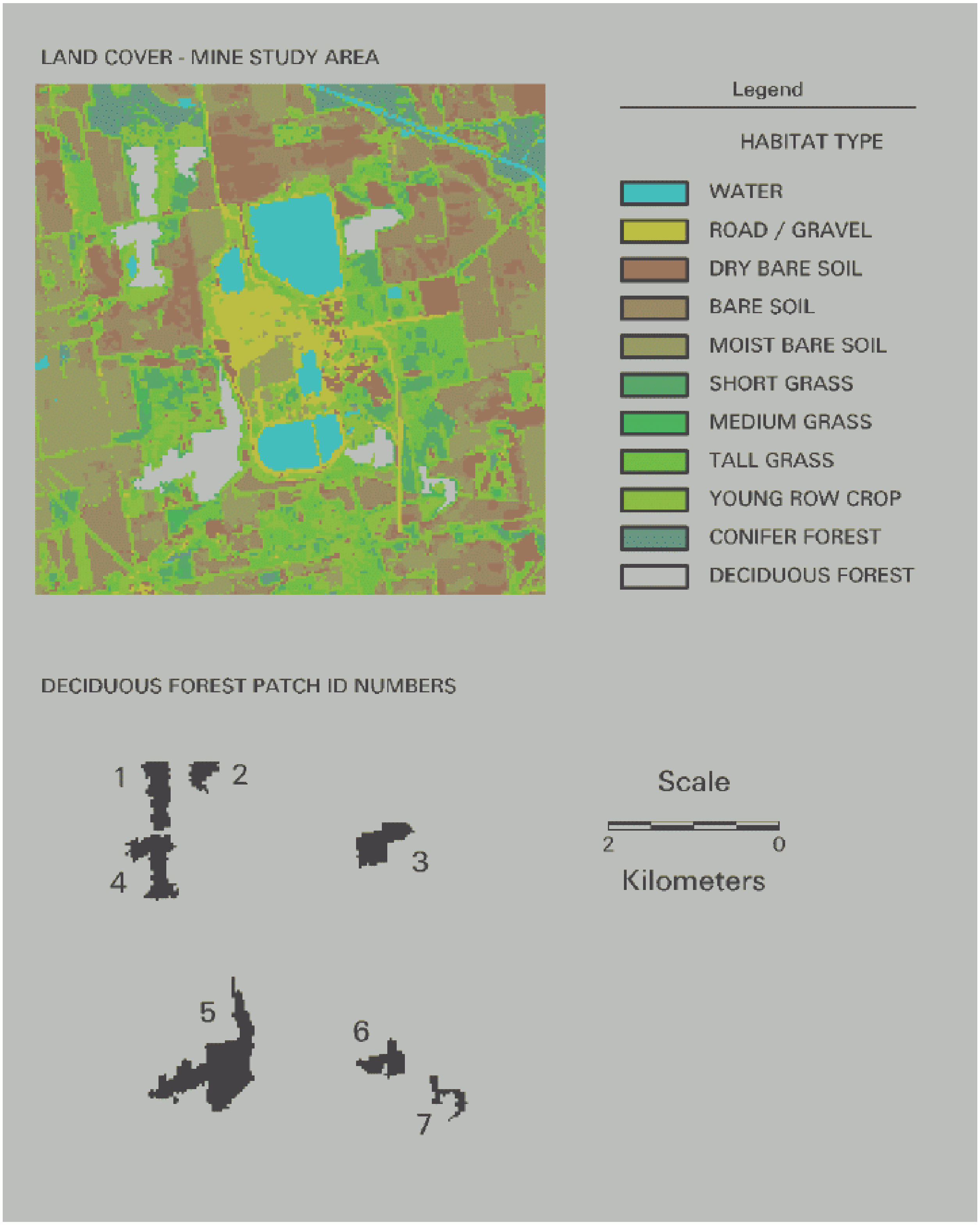

Willamette River discharge, water year 1990
Observed by USGS
Modeled by IBGC
Discharge (mm)
0
50
100
150
200
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Month of water year 1990

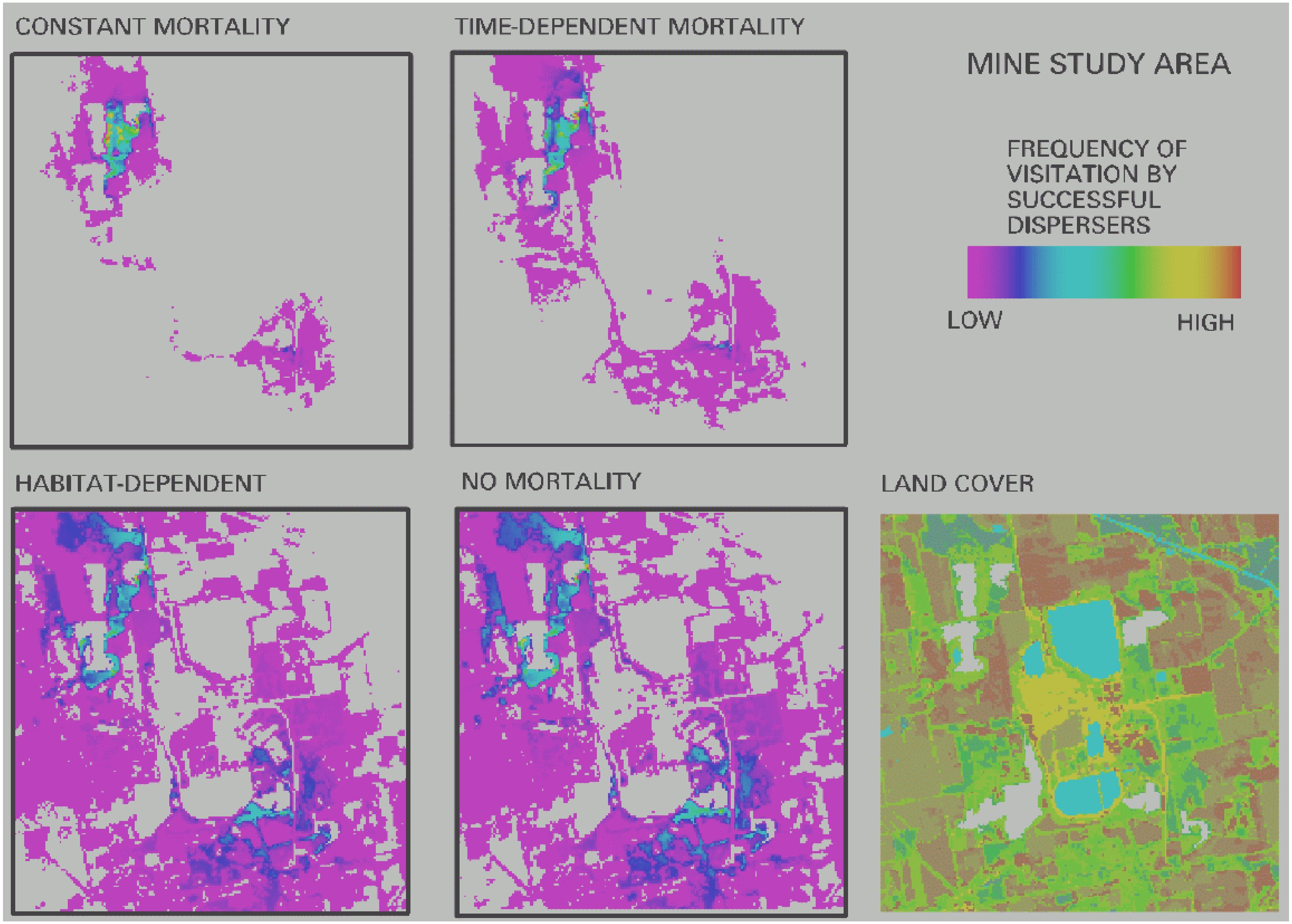

Potential net primary productivity
Units: (kg/ha)
Difference map:
1-km values minus 50-km values
50-km values greater:
-520 to -1350
-420
-320
-220
-120
0
1-km values greater:
0
100
160
260
360
460 to 810
kilometers
0
50
100

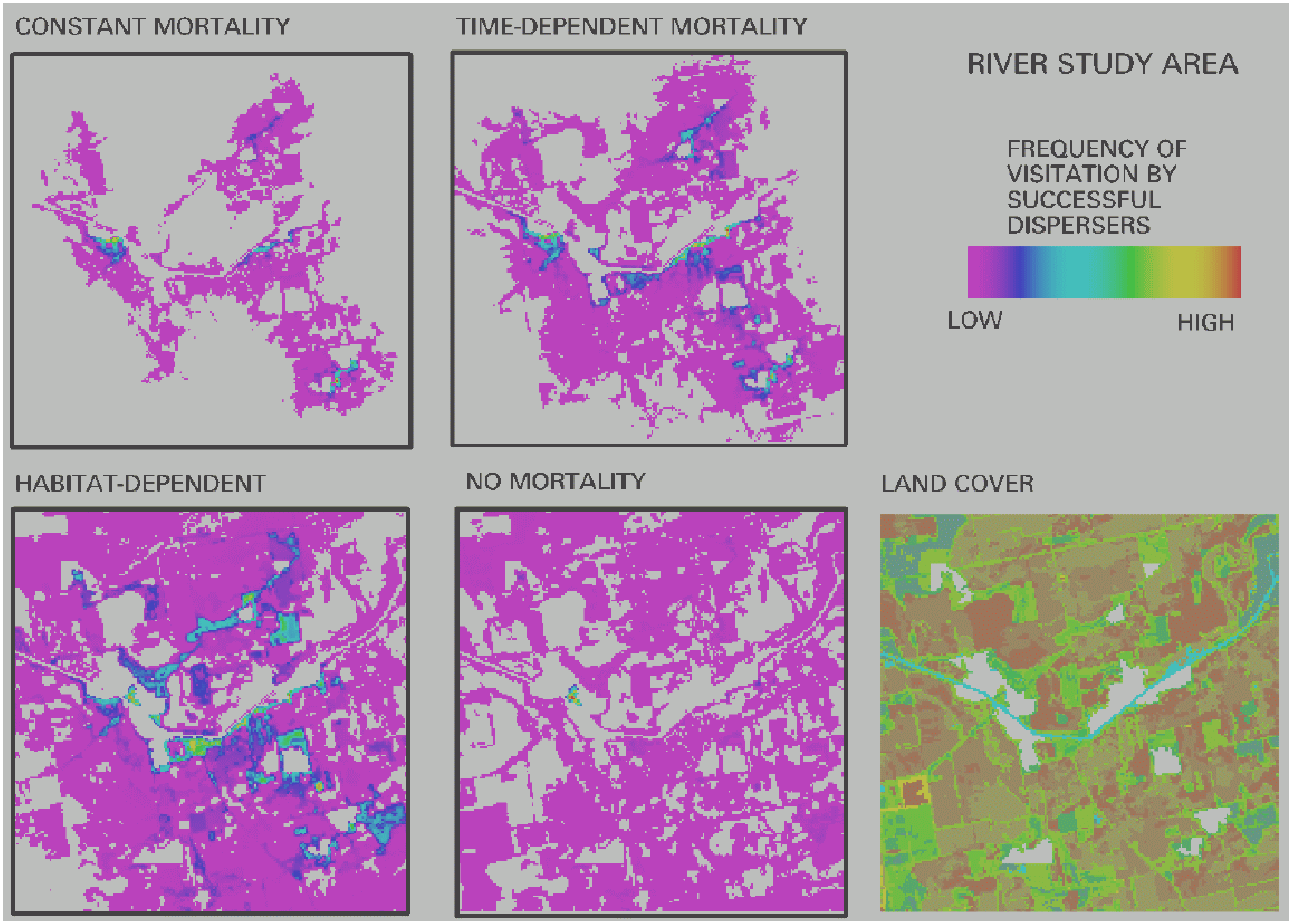

DAILY
snowpack, m ** 3 / ha
soil_h2o, m ** 3 / ha
outflow, m ** 3 / ha
trans, m ** 3 / ha
evap, m ** 3 / ha
net_C_sum, kg / ha
resp_autr, kg / ha
rain, m / ha / day
snow, m / ha / day
melt, m / ha / day
evap, m / ha / day
pot_evap, m / ha / day
net_leaf_h2o, m / ha / day
soil_h2o_frac, 0 - 1
soil_temp_frac, 0 - 1
soil_dcmp_rate, 0 - 1
lwp, -mpa
night_resp, kg / ha /day
net_C, kg / ha /day
stem_resp, kg / ha /day
root_resp, kg / ha /day
daily_C, kg / ha / day
precip, m
t_max, deg C
t_min, deg C
rel_humid, %
t_soil, deg C
rad, kj / m2 / day
day_temp, deg C
night_temp, deg C
vp_def, mb
hum_def, mic. gm / m3
daylen, sec
canopy_rad, kj / m2 / day
/files/home/towhee/dodson/_dem.shaded.relief.band0
* point_1
WAVE 0
point1 12/15/1995 09:25:00.000
1.0
0.8
0.6
0.4
0.2
0.0
2.0x10^4
1.5x10^4
1.0x10^4
5.0x10^3
0
−5.0x10^3
0
200
day
soil_h2o_frac, 0 — 1
net_C_sum, kg / ha

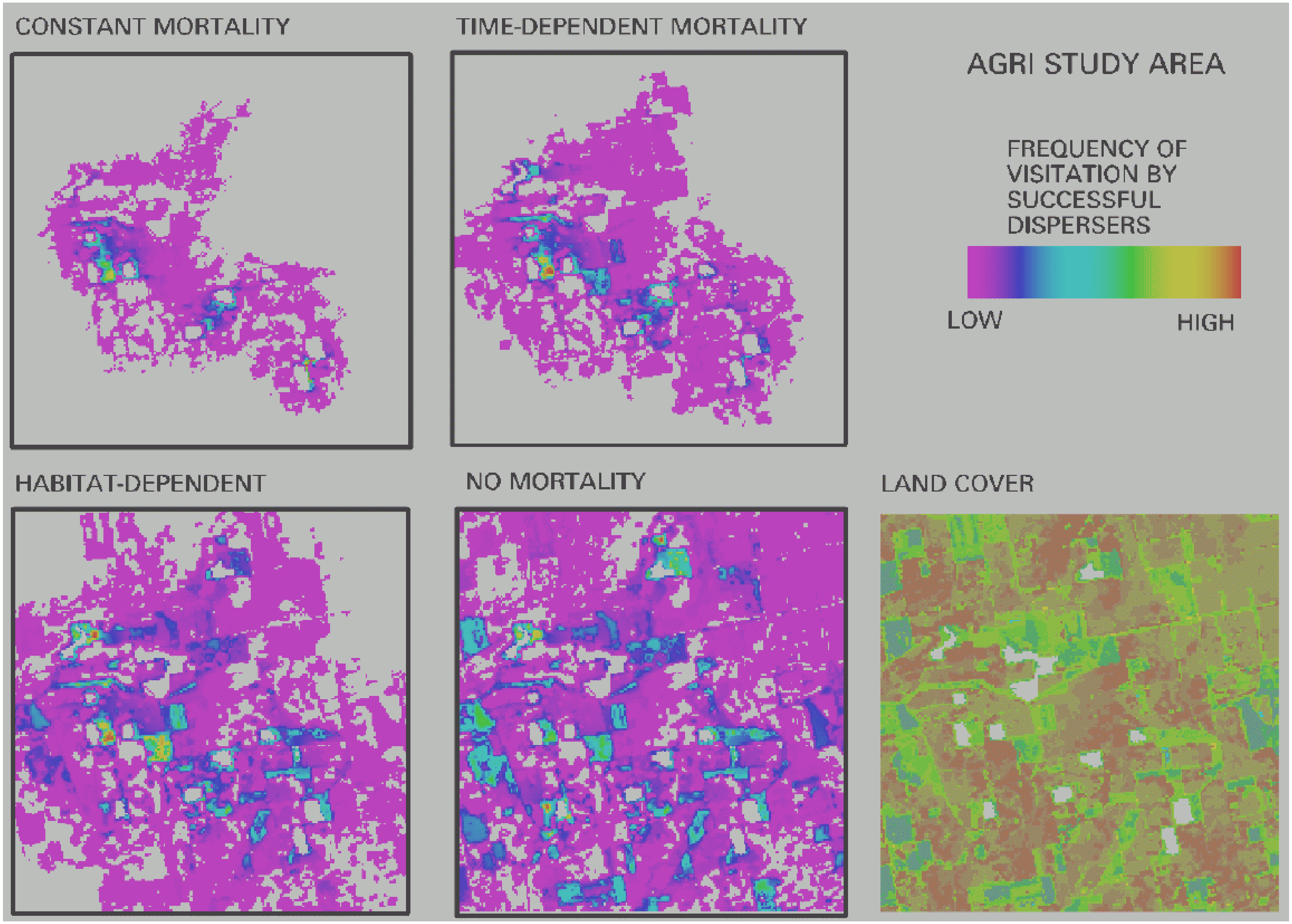

DAILY
snowpack, m ** 3 / ha
soil_h2o, m ** 3 / ha
outflow, m ** 3 / ha
trans, m ** 3 / ha
evap, m ** 3 / ha
net_C_sum, kg / ha
resp_autr, kg / ha
rain, m / ha / day
snow, m / ha / day
melt, m / ha / day
evap, m / ha / day
pot_evap, m / ha / day
net_leaf_h2o, m / ha / day
soil_h2o_frac, 0 - 1
soil_temp_frac, 0 - 1
soil_dcmp_rate, 0 - 1
lwp, -mpa
lc_pdr, m / sec
lc_nmr
lc_hdr
lc_rr
lc, m / sec
pm_trans, m / lai / day
trans, m / ha / day
outflow, m / ha / day
nit_coeff
light_coeff, m / sec
temp_coeff
meso_cond, m / sec
gross_psn, kg / lai
daily_psn, kg / ha /day
night_resp, kg / ha /day
net_C, kg / ha /day
stem_resp, kg / ha /day
root_resp, kg / ha /day
daily_C, kg / ha / day
precip, m
t_max, deg C
t_min, deg C
rel_humid, %
t_soil, deg C
rad, kj / m2 / day
day_temp, deg C
night_temp, deg C
vp_def, mb
hum_def, mic. gm / m3
daylen, sec
canopy_rad, kj / m2 / day
soil_dcmp_rate, kg/ha
litter_C_dcmp, kg/ha
litter_C_atm, kg/ha
rootlitter_C_dcmp, kg/ha
rootlitter_C_atm, kg/ha
cwd_C_dcmp, kg/ha
cwd_C_atm, kg/ha
soil_C_dcmp, kg/ha
soil_C_atm, kg/ha
total_C_atm, kg/ha
daily_nep, kg/ha
-- PLOT --
-- Return to prev. menu --
/files/home/towhee/dodson/_dem.shaded.relief.band0
* point 1
* point 2
* point 3
WAVE 0
995 09:45:00.000
2.0×10^4
1.5×10^4
1.0×10^4
5.0×10^3
0
−5.0×10^3
0
200
day
point1: net_C_sum, kg / ha

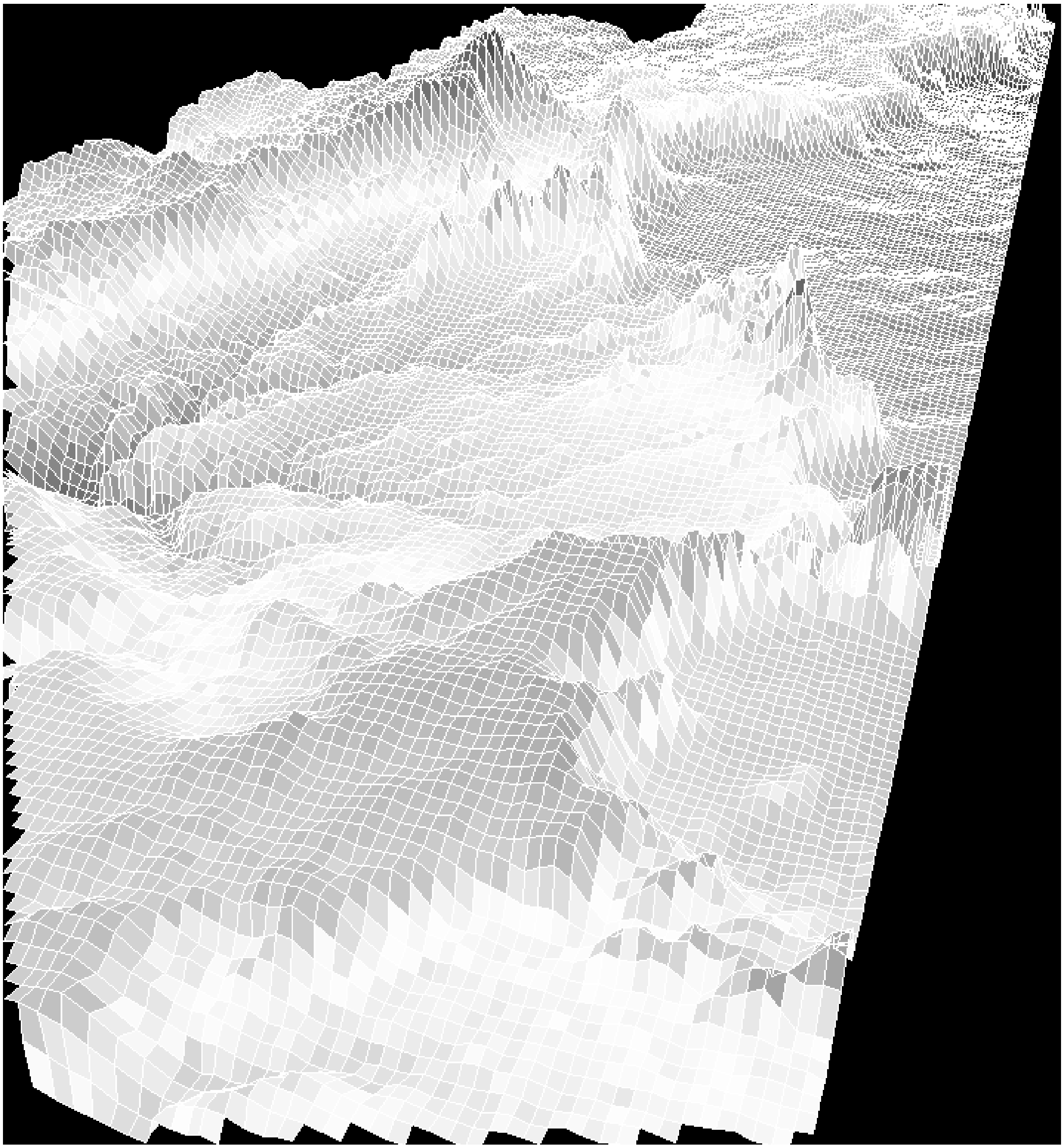

point3: net_C_sum, kg / ha

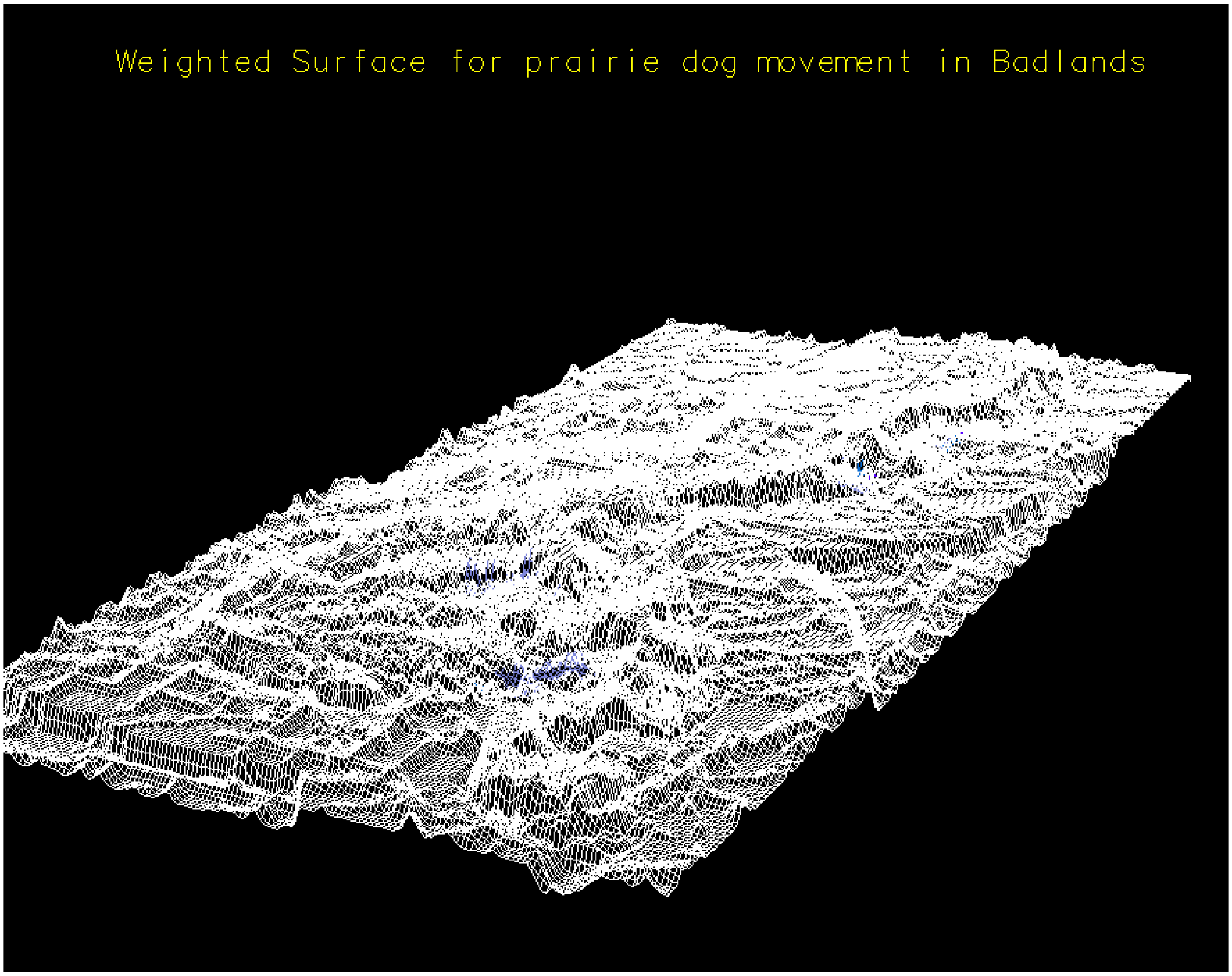

LAND COVER - MINE STUDY AREA
Legend
HABITAT TYPE
WATER
ROAD / GRAVEL
DRY BARE SOIL
BARE SOIL
MOIST BARE SOIL
SHORT GRASS
MEDIUM GRASS
TALL GRASS
YOUNG ROW CROP
CONIFER FOREST
DECIDUOUS FOREST
DECIDUOUS FOREST PATCH ID NUMBERS
1
2
3
4
5
6
7
Scale
2
0
Kilometers

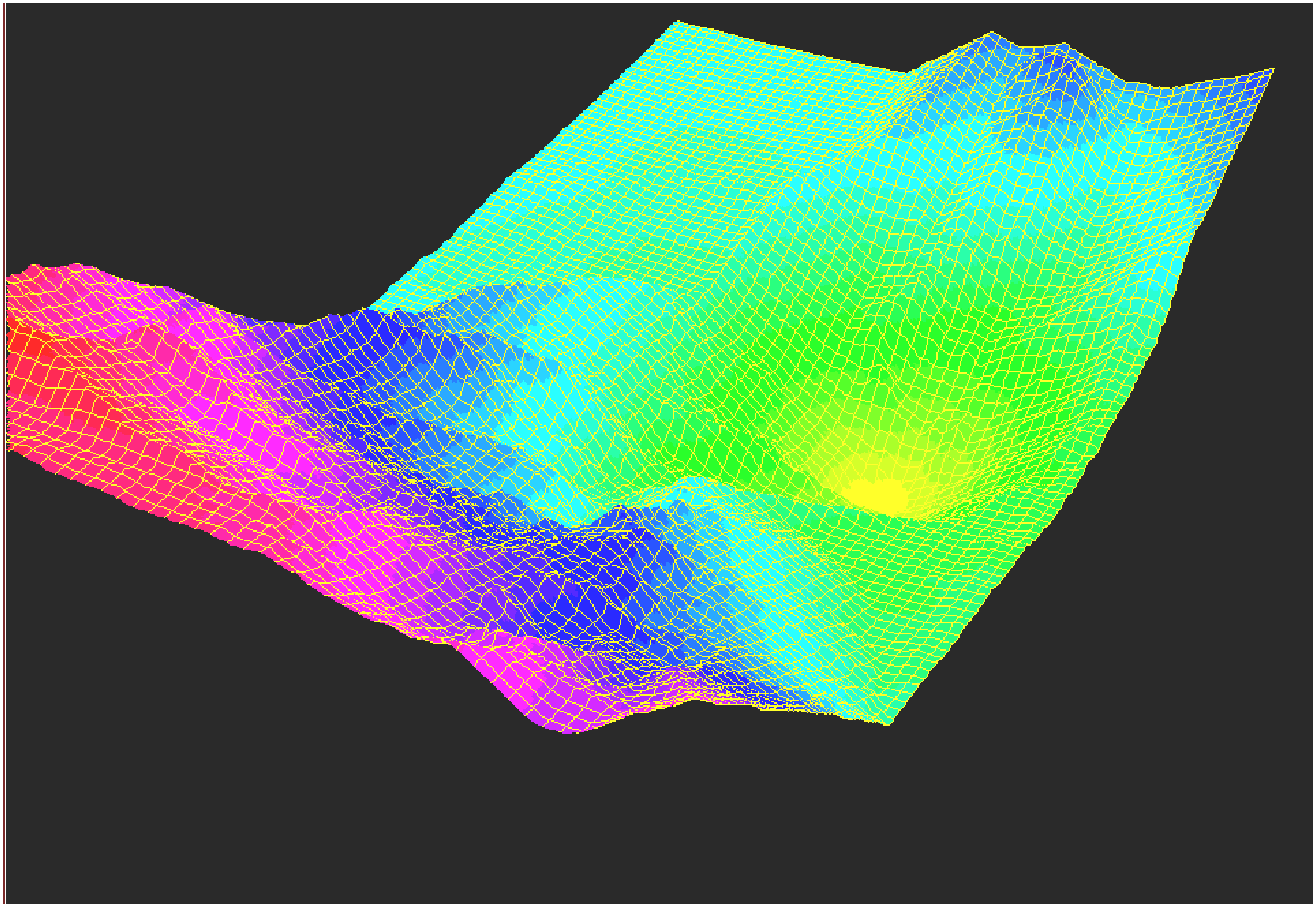

CONSTANT MORTALITY
TIME-DEPENDENT MORTALITY
MINE STUDY AREA
FREQUENCY OF VISITATION BY SUCCESSFUL DISPERSERS
LOW
HIGH
HABITAT-DEPENDENT
NO MORTALITY
LAND COVER

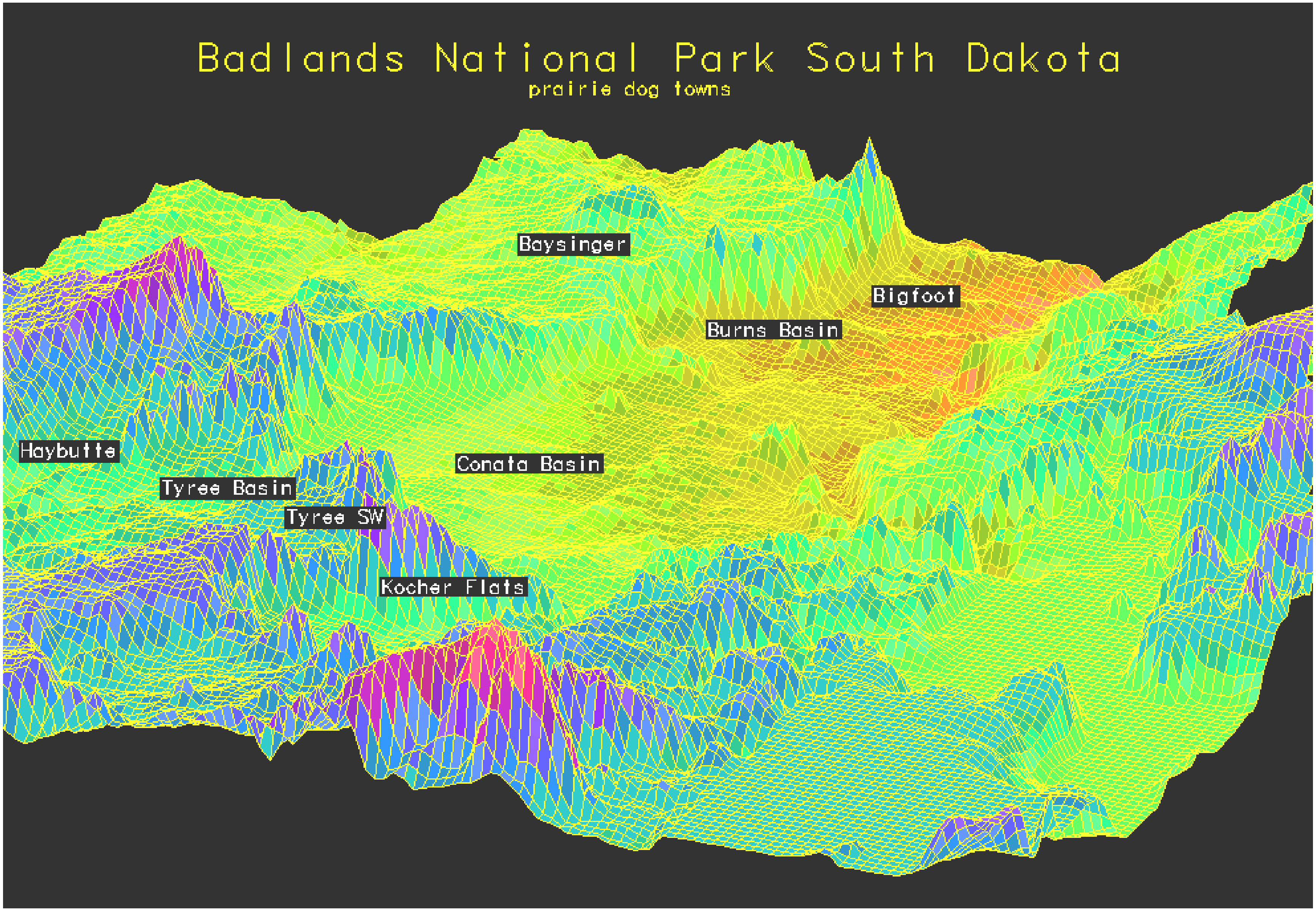

CONSTANT MORTALITY
TIME-DEPENDENT MORTALITY
RIVER STUDY AREA
FREQUENCY OF
VISITATION BY
SUCCESSFUL
DISPERSERS
LOW
HIGH
HABITAT-DEPENDENT
NO MORTALITY
LAND COVER

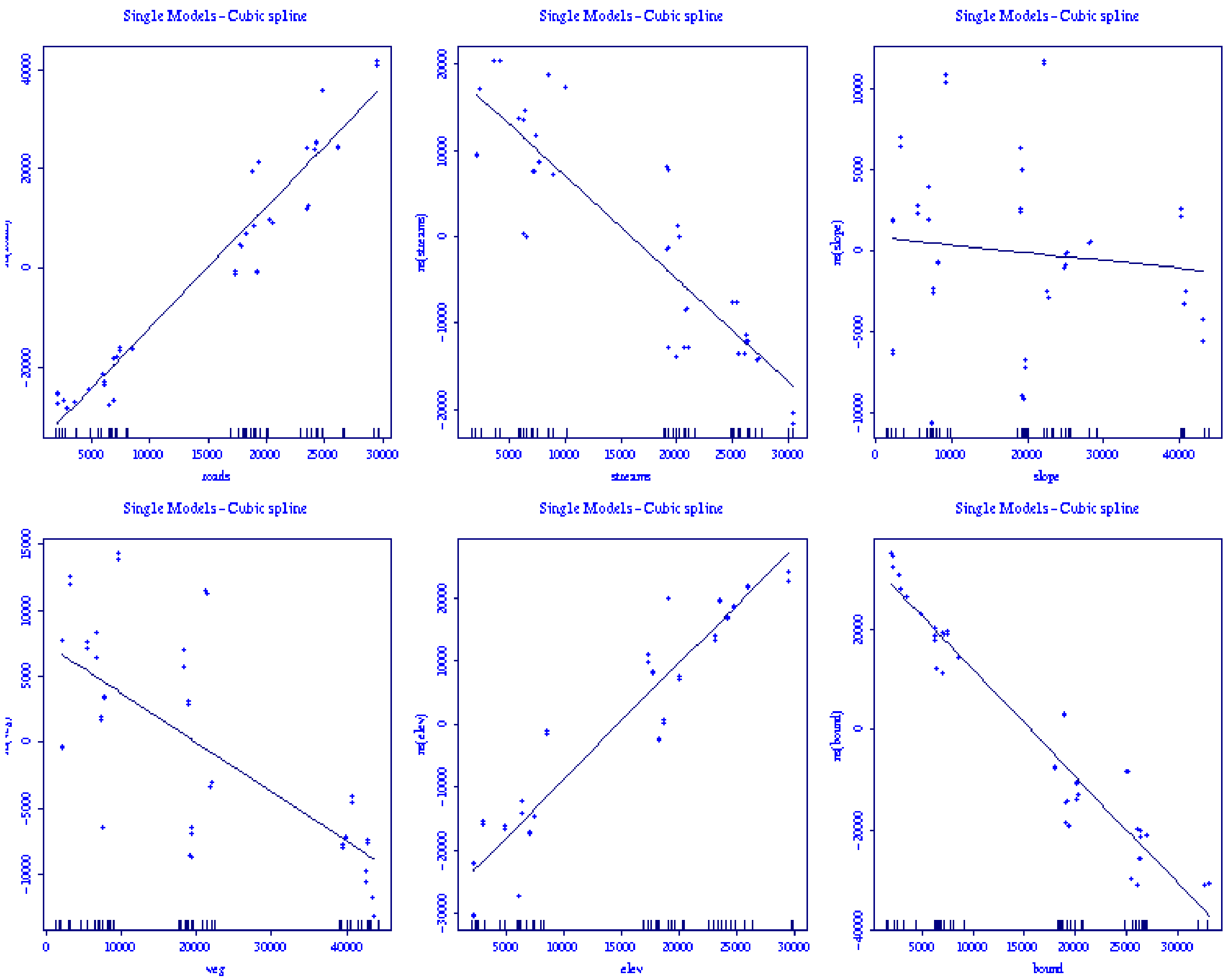

CONSTANT MORTALITY
TIME-DEPENDENT MORTALITY
AGRI STUDY AREA
FREQUENCY OF
VISITATION BY
SUCCESSFUL
DISPERSERS
LOW
HIGH
HABITAT-DEPENDENT
NO MORTALITY
LAND COVER

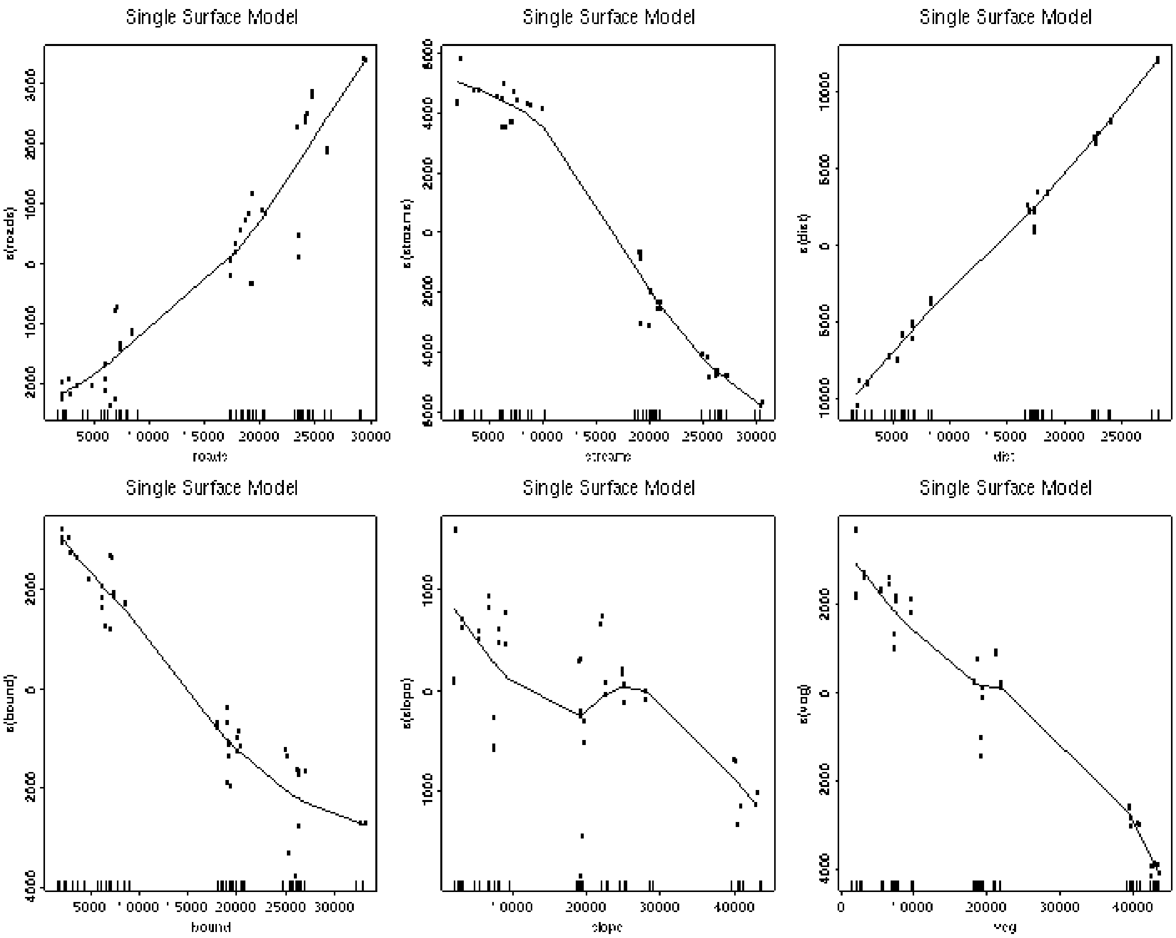


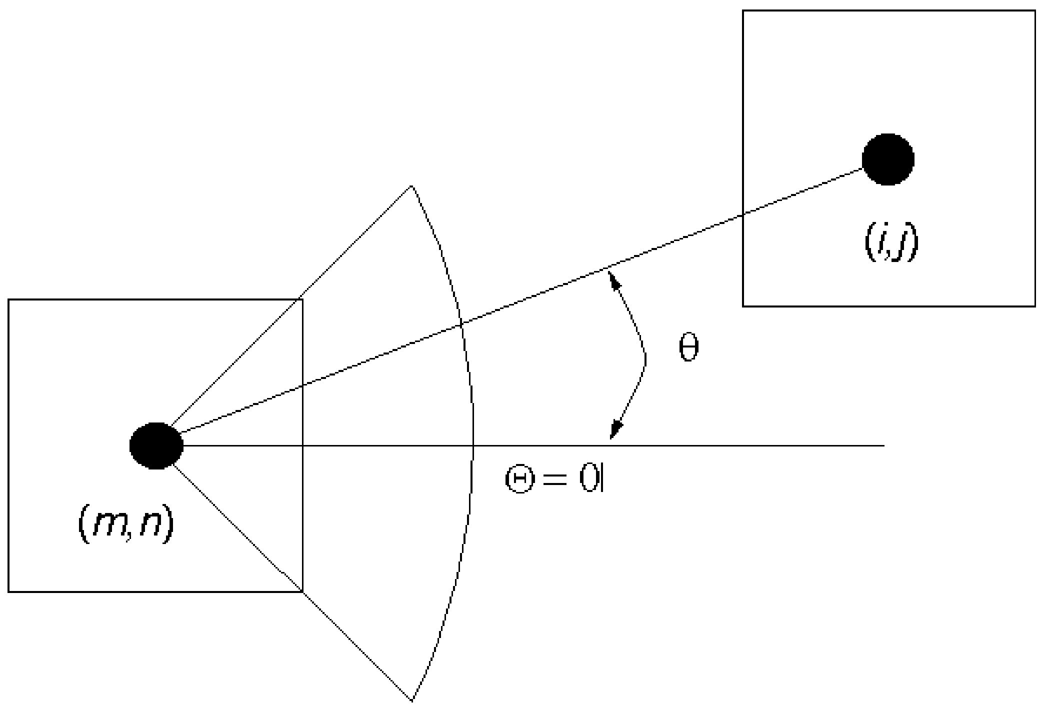

Weighted Surface for prairie dog movement in Badlands



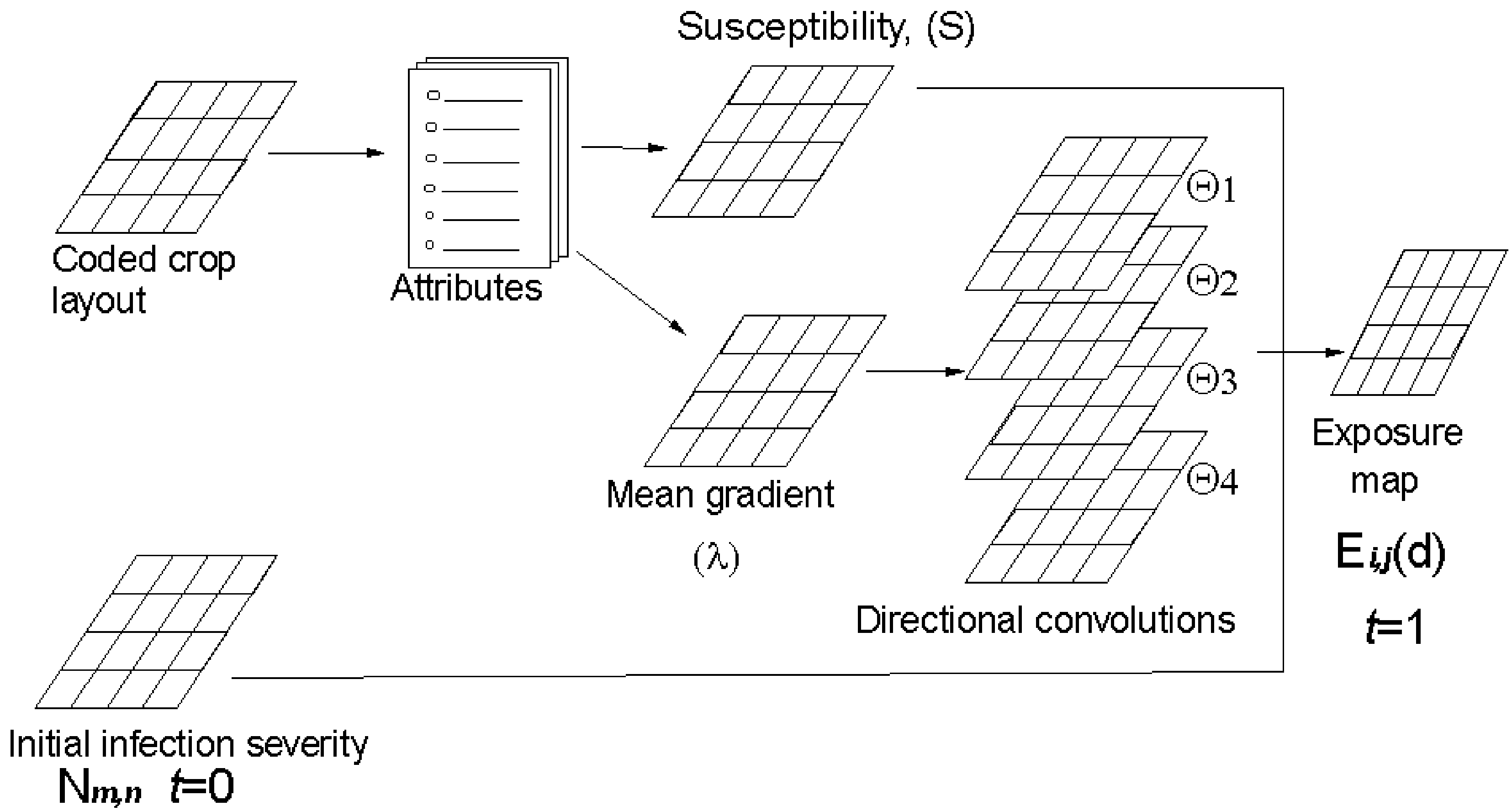

Badlands National Park South Dakota
prairie dog towns
Baysinger
Bigfoot
Burns Basin
Haybutte
Conata Basin
Tyree Basin
Tyree SW
Kocher Flats

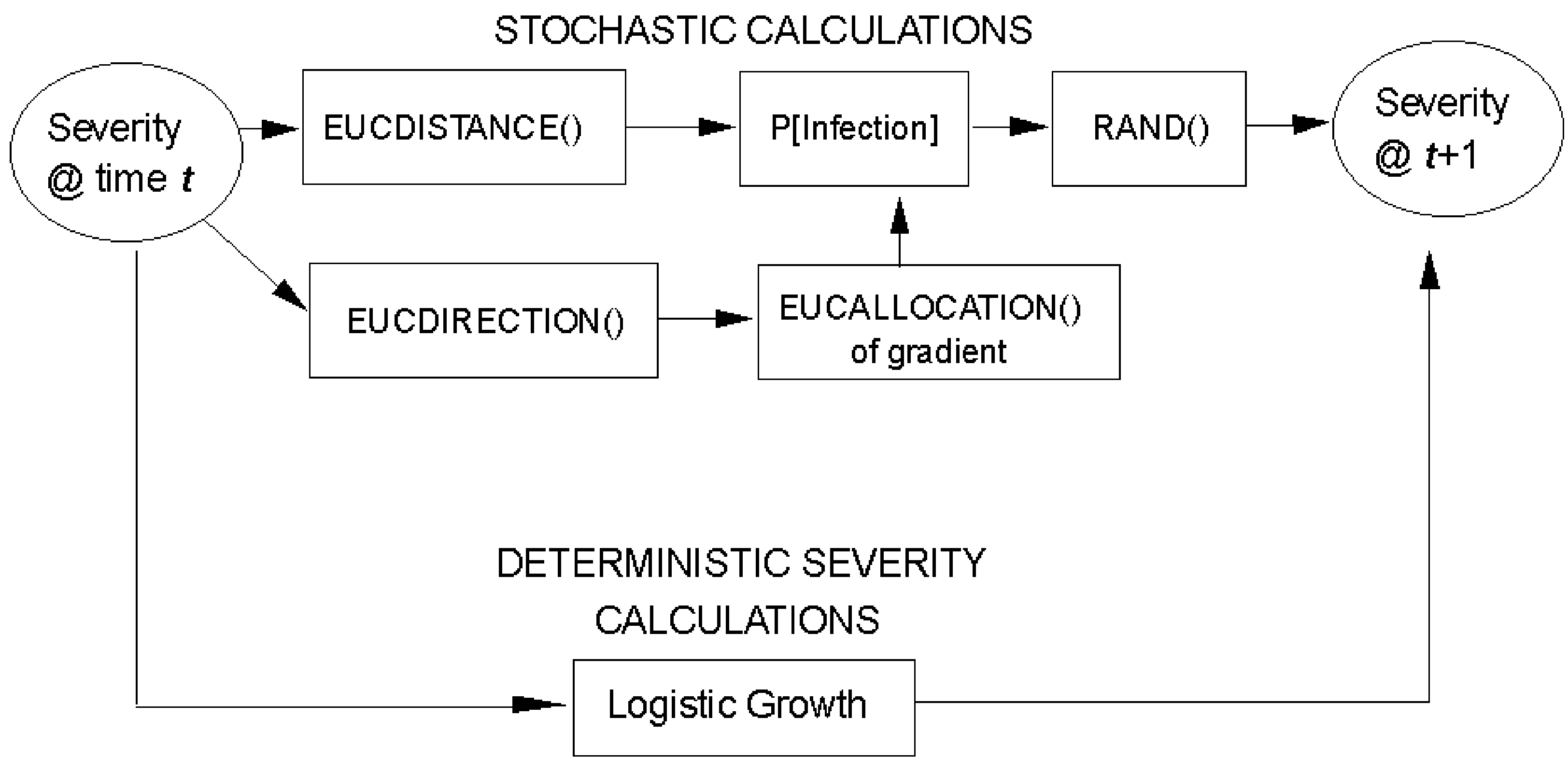

Single Models - Cubic spline
roads
Single Models - Cubic spline
m(streams)
streams
Single Models - Cubic spline
m(slope)
slope
Single Models - Cubic spline
veg
Single Models - Cubic spline
m(elev)
elev
Single Models - Cubic spline
m(bound)
bound

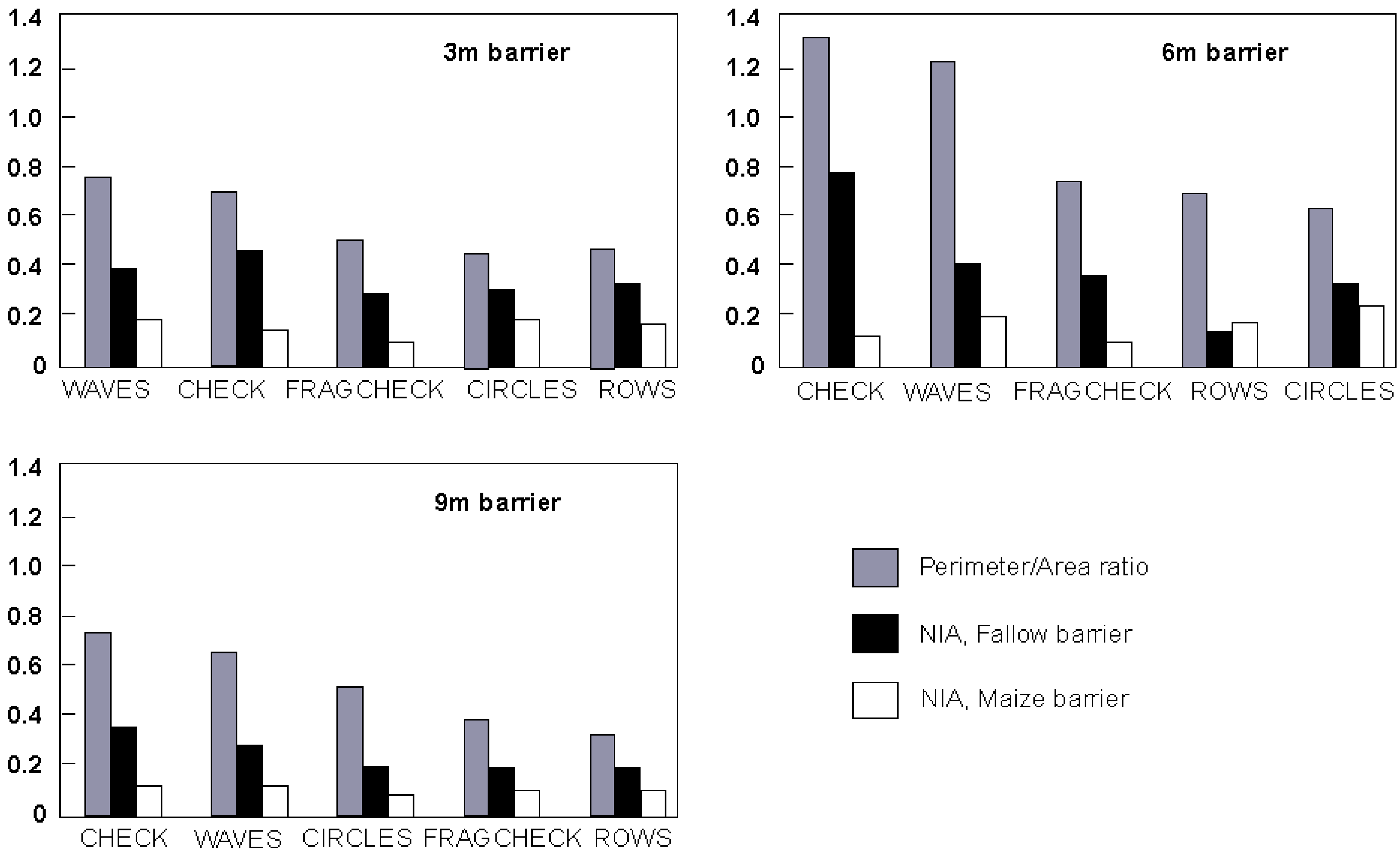

Single Surface Model
s(roads)
roads
Single Surface Model
s(streams)
streams
Single Surface Model
s(dist)
dist
Single Surface Model
s(bound)
bound
Single Surface Model
s(slope)
slope
Single Surface Model
s(veg)
veg

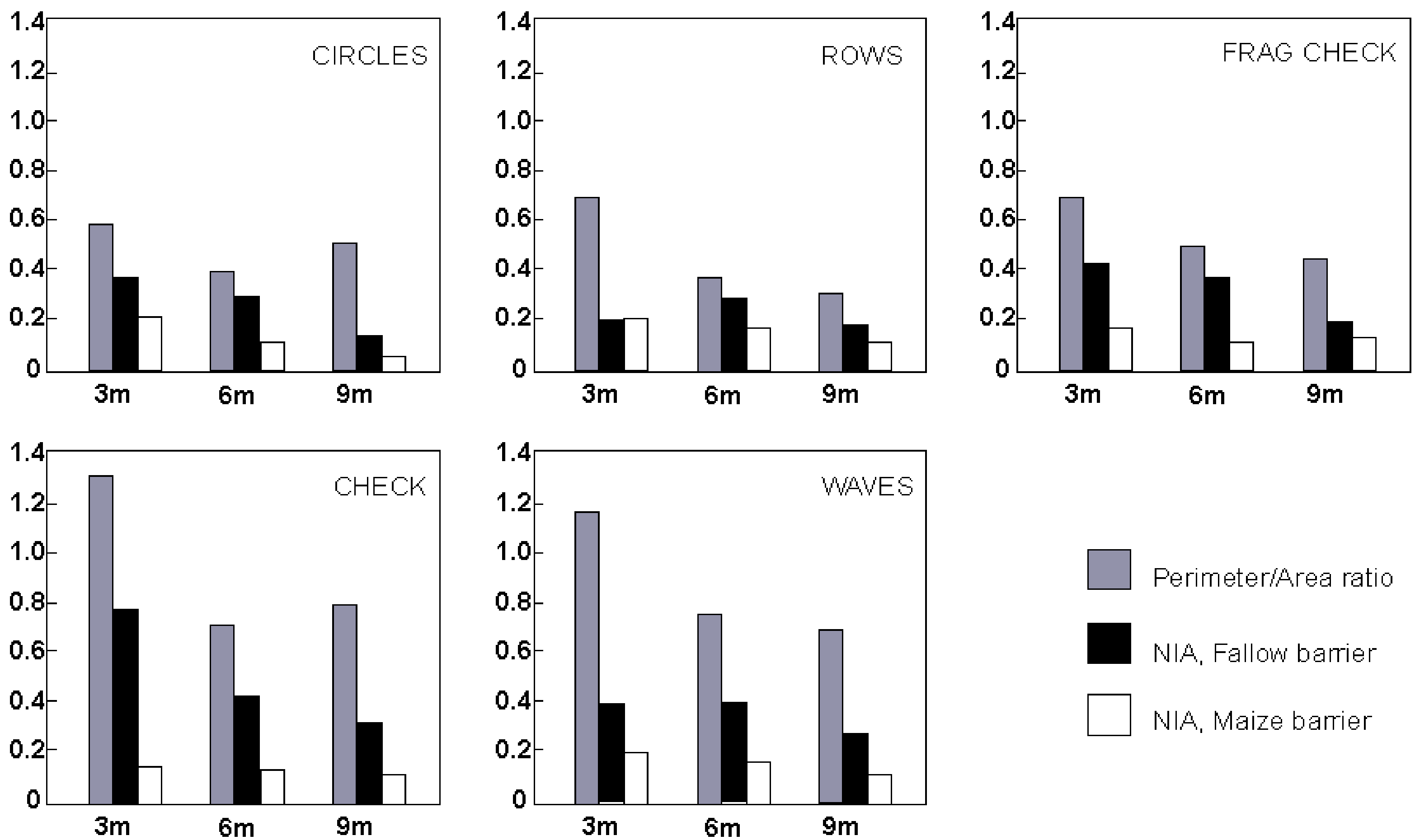


$$E_{i,j}(\mathrm{d}) = \lambda_{m,n,\Theta} e^{-\lambda_{m,n,\Theta} \mathrm{d}}$$

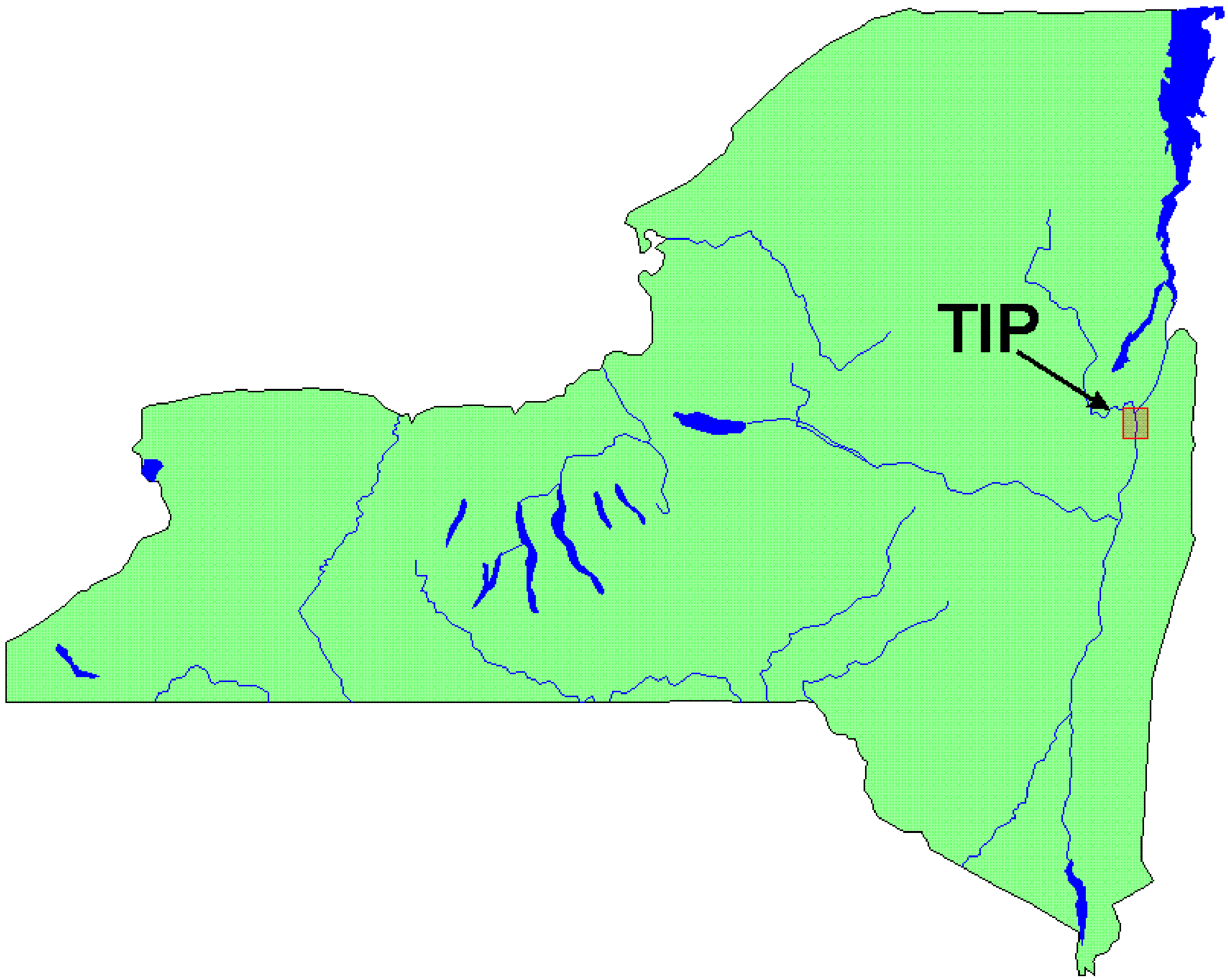

Susceptibility, (S)
Coded crop layout
Attributes
Θ1
Θ2
Θ3
Θ4
Exposure map
Mean gradient
(λ)
$E_{i,j}(d)$
$t=1$
Directional convolutions
Initial infection severity
$N_{m,n}$ $t=0$


STOCHASTIC CALCULATIONS
Severity @ time t
EUCDISTANCE()
P[Infection]
RAND()
Severity @ t+1
EUCDIRECTION()
EUCALLOCATION() of gradient
DETERMINISTIC SEVERITY CALCULATIONS
Logistic Growth

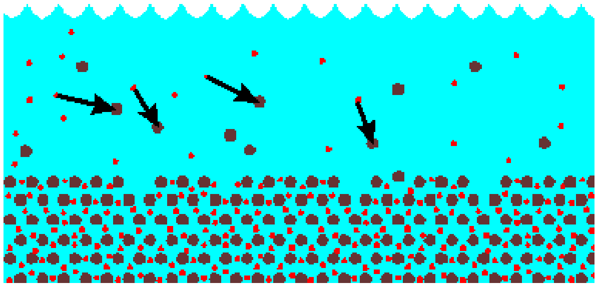

3m barrier
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
WAVES
CHECK
FRAGCHECK
CIRCLES
ROWS
6m barrier
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
CHECK
WAVES
FRAGCHECK
ROWS
CIRCLES
9m barrier
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
CHECK
WAVES
CIRCLES
FRAGCHECK
ROWS
Perimeter/Area ratio
NIA, Fallow barrier
NIA, Maize barrier



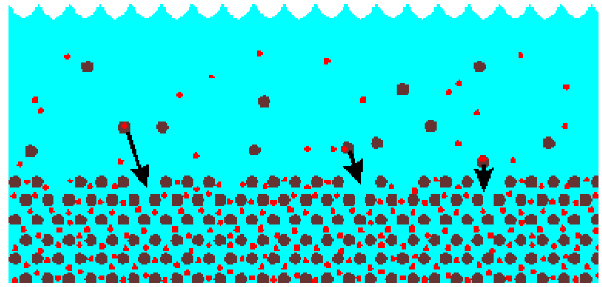
CIRCLES
ROWS
FRAG CHECK
CHECK
WAVES
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
3m
6m
9m
Perimeter/Area ratio
NIA, Fallow barrier
NIA, Maize barrier


TIP

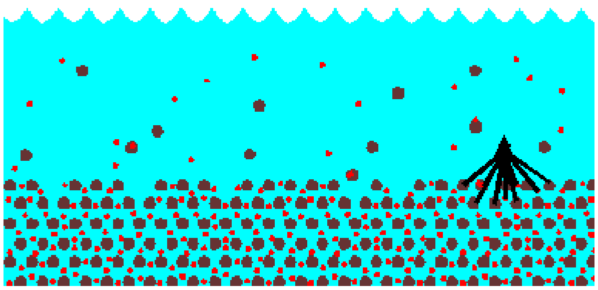
Deposition and resuspension



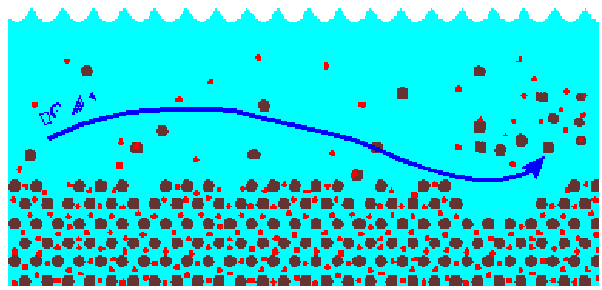
Partitioning: material attaches to suspended solids in water.



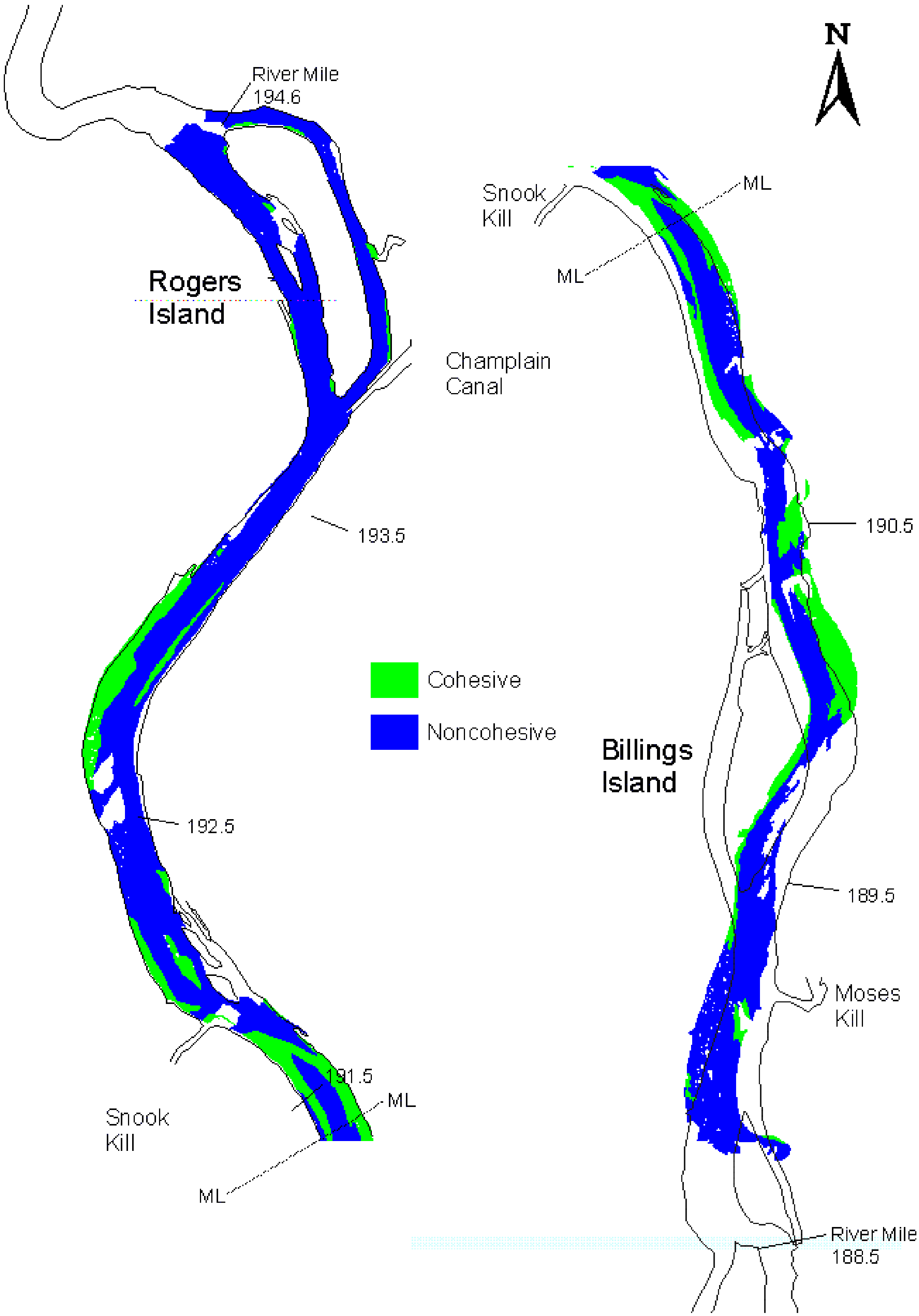
Settling - solids with attached materials fall to the bottom.

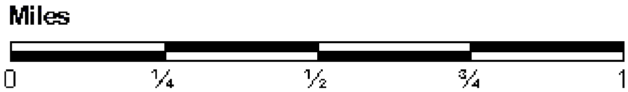


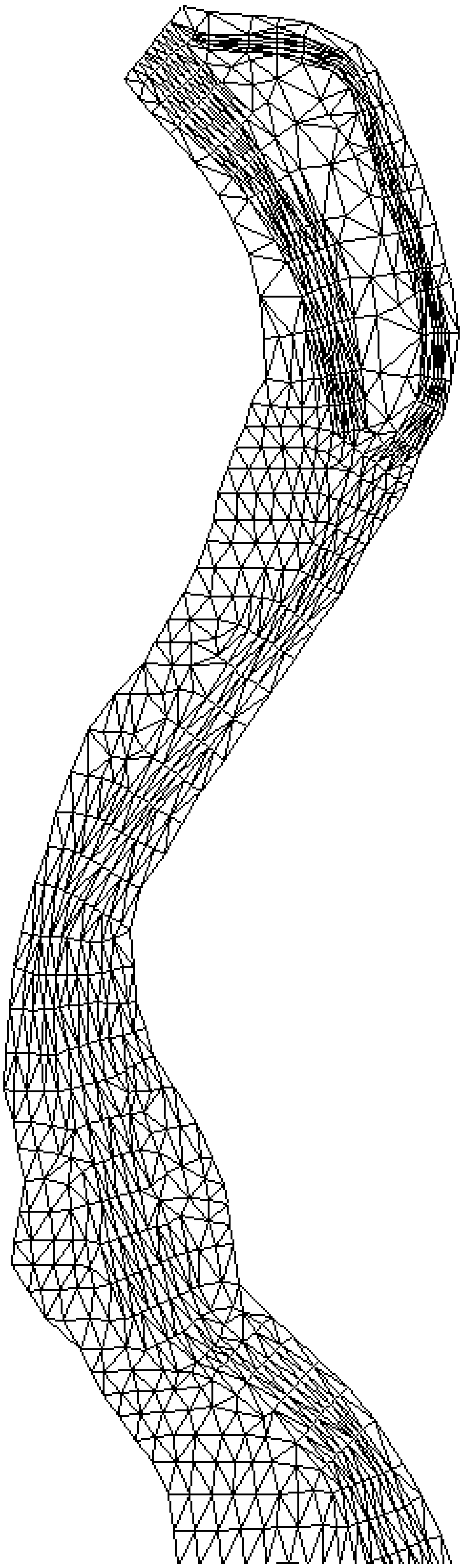
Resuspension - high flow remobilizes sediments and attached materials.

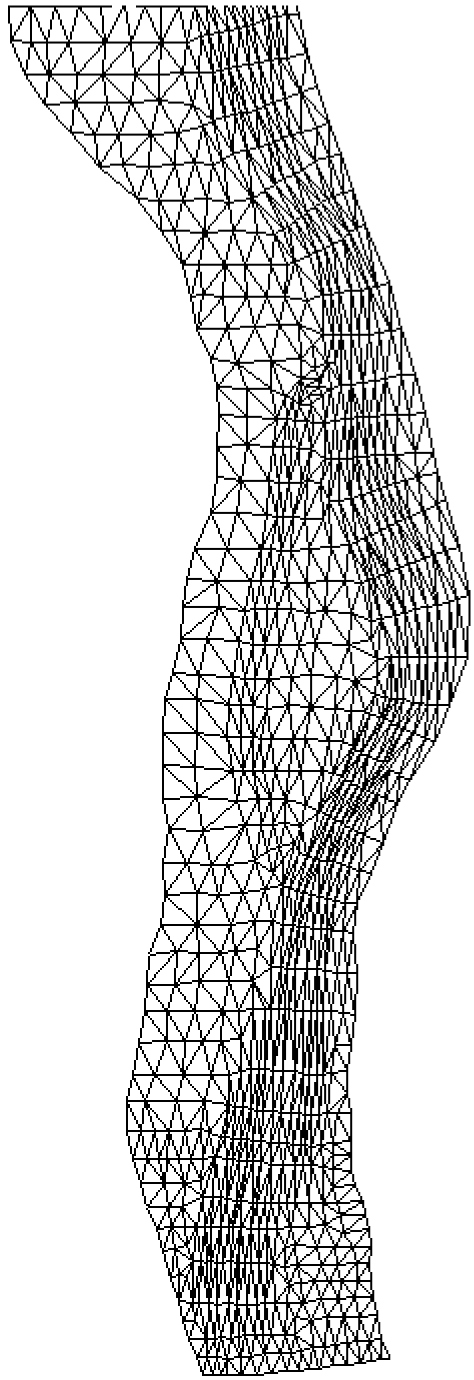


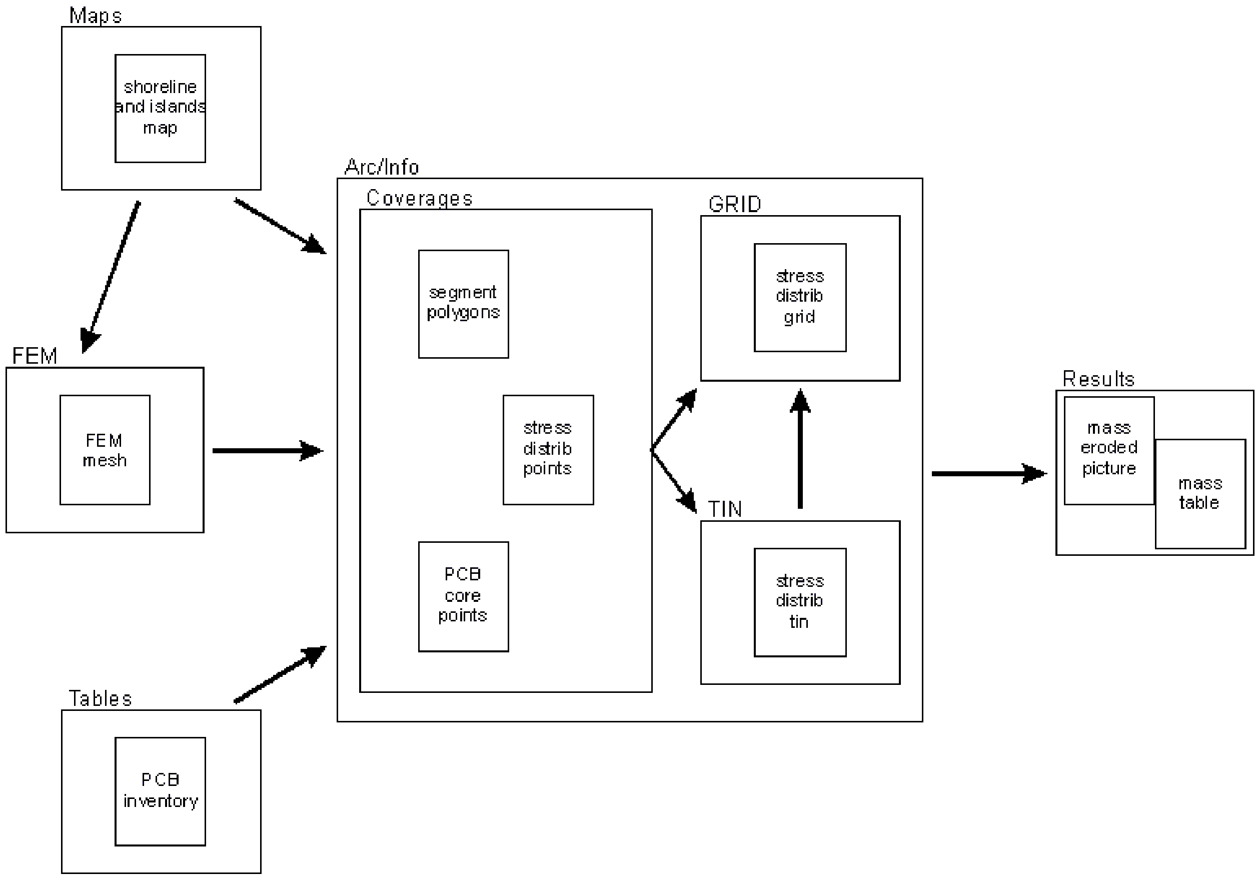
Resuspended materials separate from solids back into the water.


N
River Mile
194.6
Rogers
Island
Champlain
Canal
193.5
Cohesive
Noncohesive
192.5
191.5
ML
Snook
Kill
ML
Snook
Kill
ML
ML
190.5
Billings
Island
189.5
Moses
Kill
River Mile
188.5


Miles
0
¼
½
¾
1

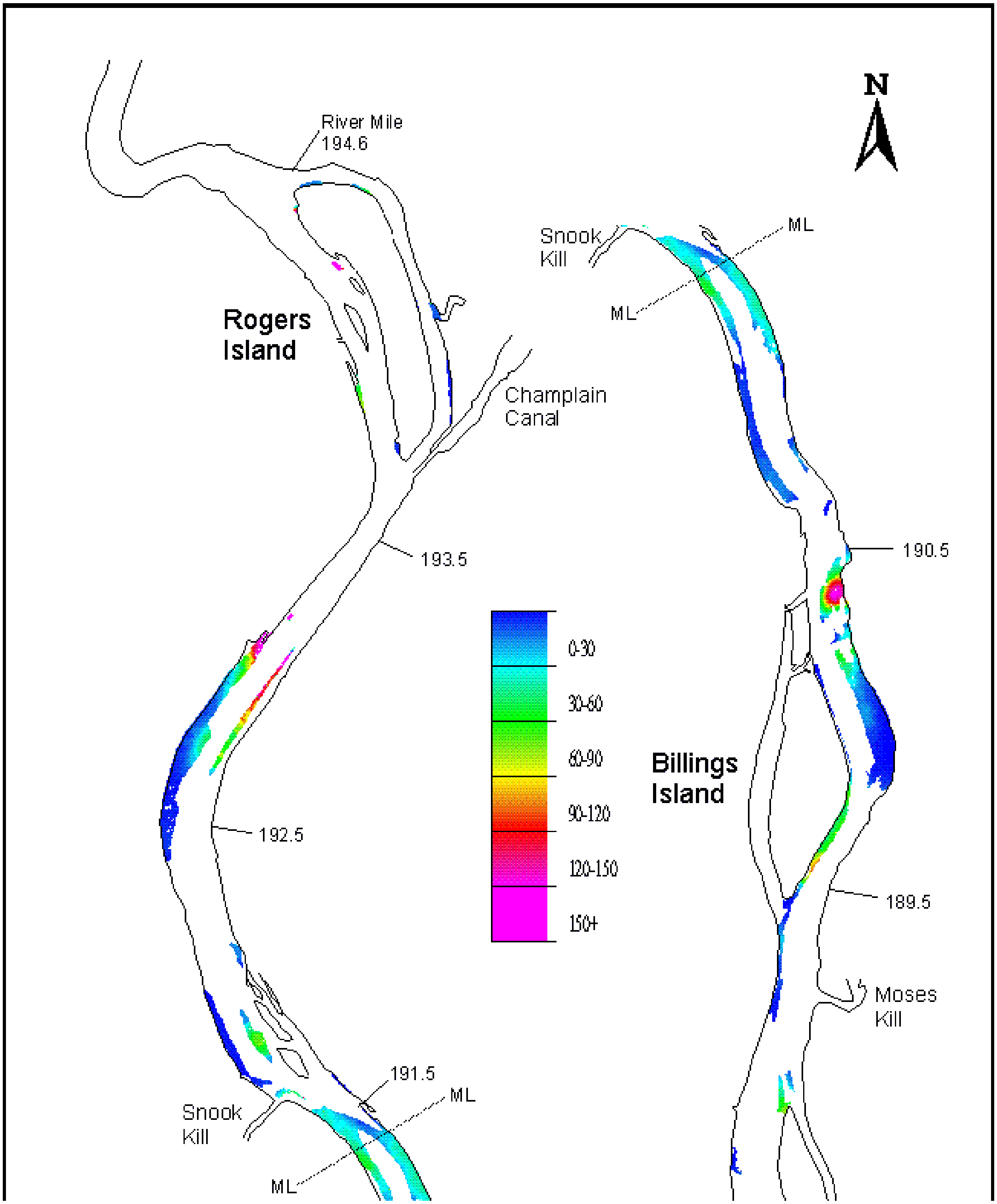


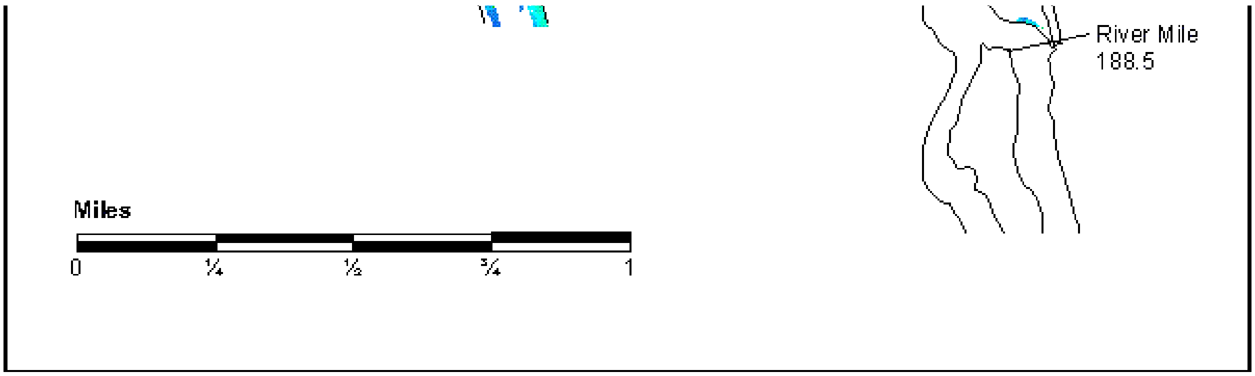



Maps
shoreline and islands map
FEM
FEM mesh
Tables
PCB inventory
Arc/Info
Coverages
segment polygons
stress distrib points
PCB core points
GRID
stress distrib grid
TIN
stress distrib tin
Results
mass eroded picture
mass table

Mass eroded (kg/s), cohesive sediments

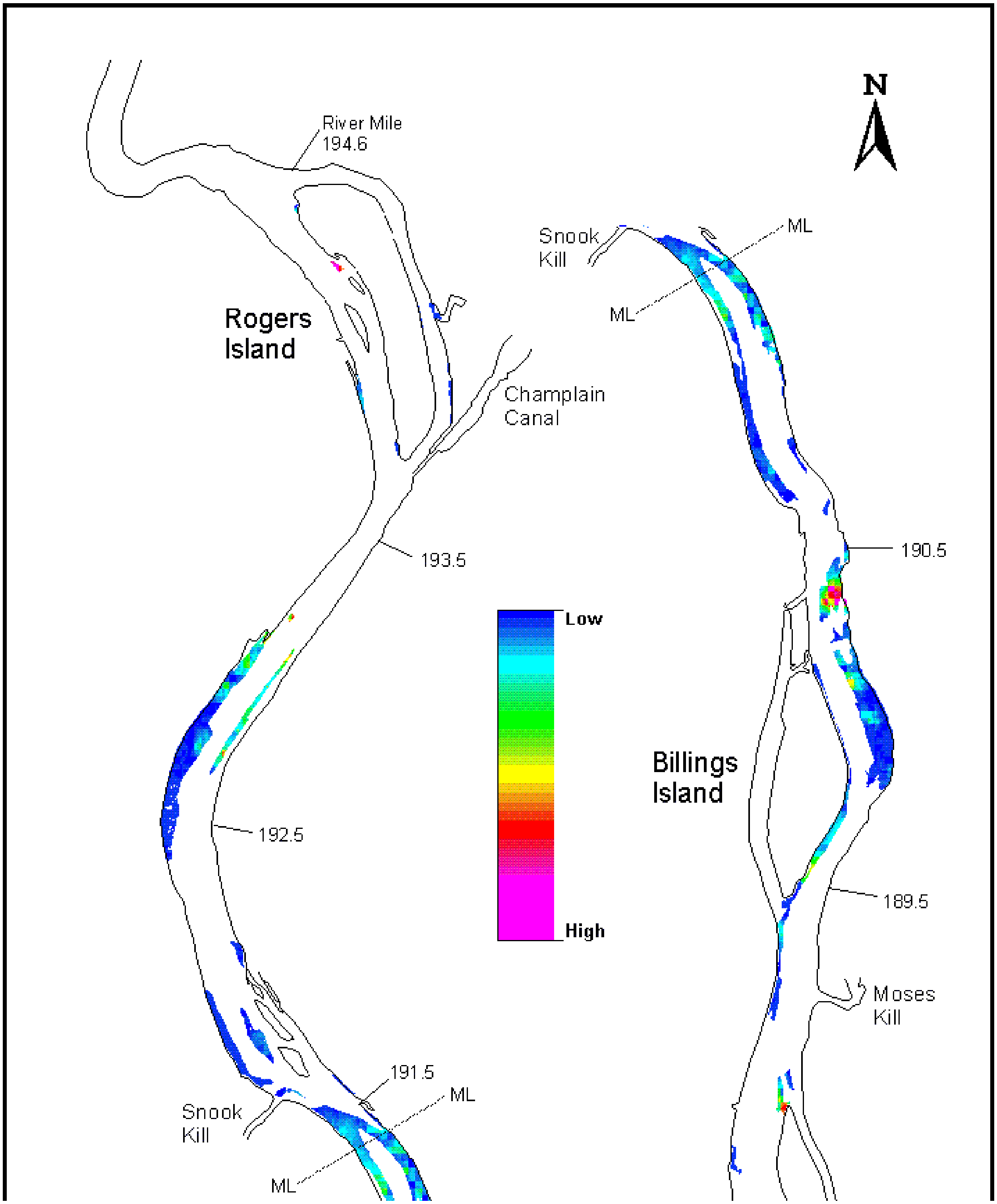
Thompson Island Pool, 100-year event (Q = 47,300 cfs)

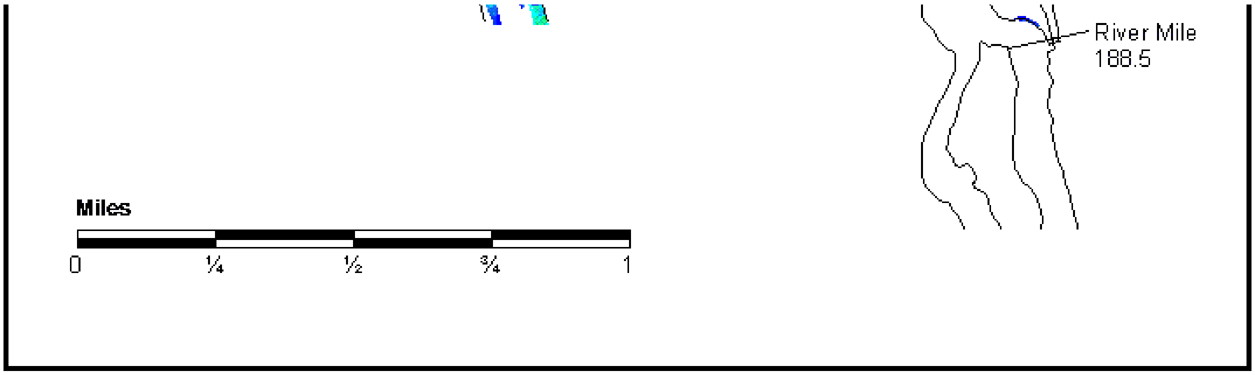


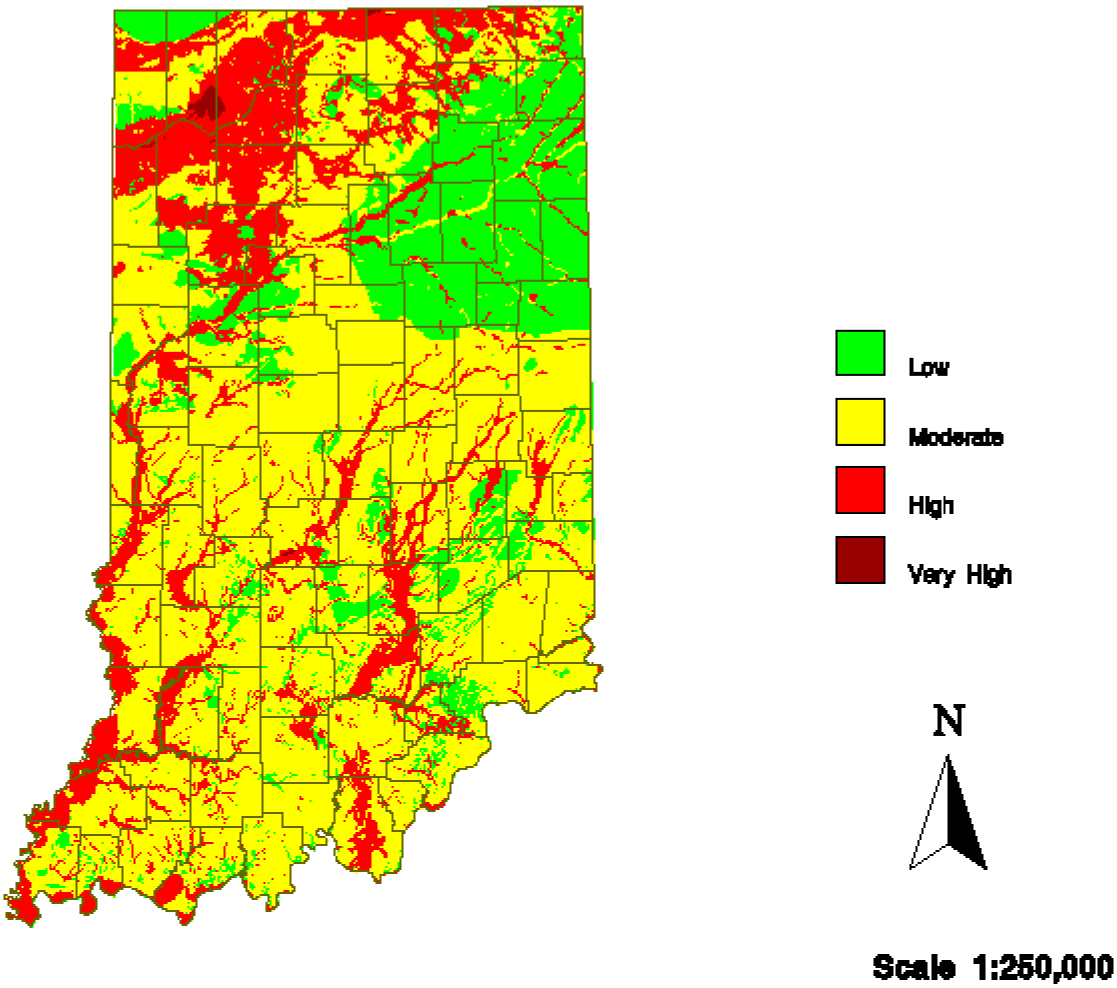

River Mile
188.5
Miles
0
¼
½
¾
1

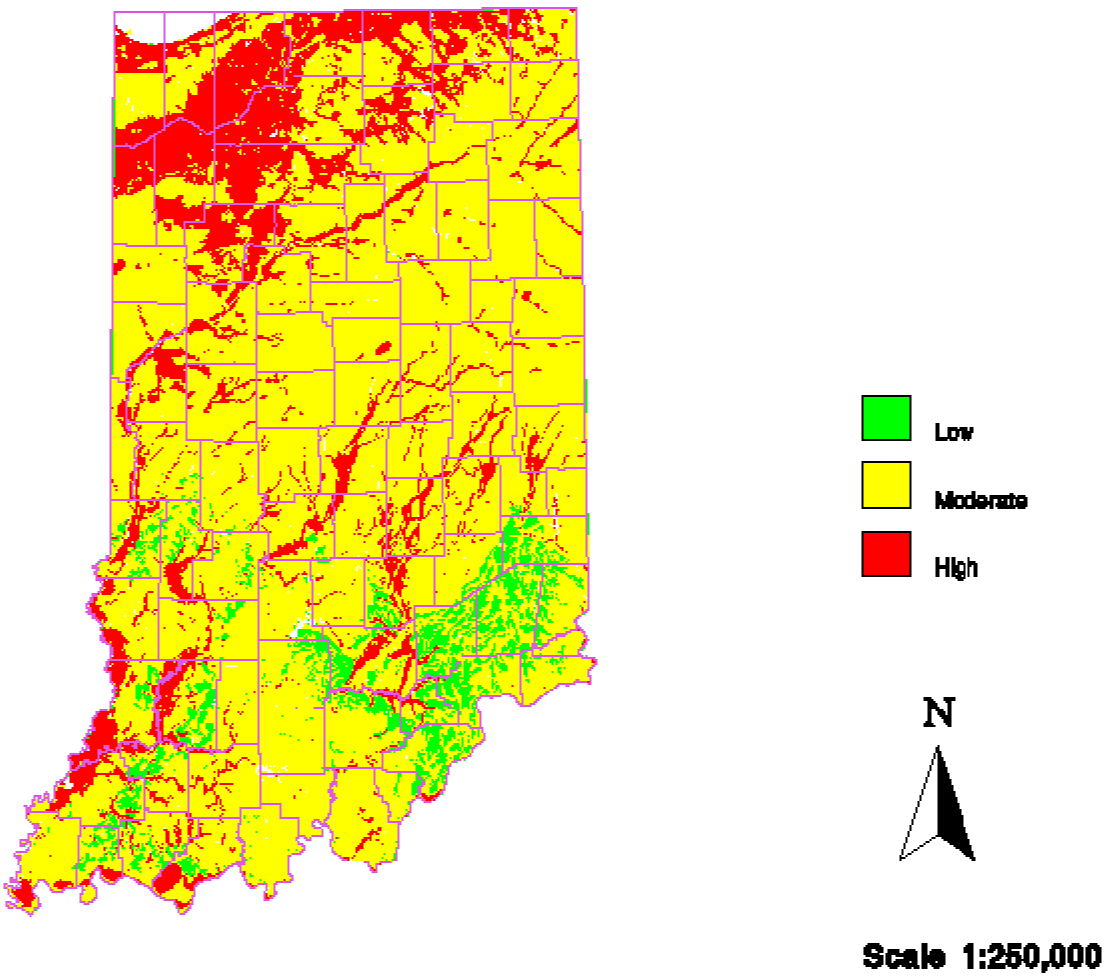
PCBs eroded, cohesive sediments

Thompson Island Pool, 100-year event (Q = 47,300 cfs)




River Mile
188.5
Miles
0
¼
½
¾
1


Low
Moderate
High
Very High
N
Scale 1:250,000


Low
Moderate
High
N
Scale 1:250,000

Jochen H. Albrecht

Universal GIS Operations for Environmental Modeling

In the first part of this paper, the author presents a set of 20 universal analytical GIS operations, i.e. operations that are independent of data structures, yet cover the full range of analytical capabilities of current vendor-based GIS. This set represents a user's task-oriented view of GIS functionality rather than that of a technically oriented developer. Their function is readily apprehended by any spatially aware individual and does not require any knowledge about abstract spatial concepts. These operations constitute the foundation of a shell that has been developed at the Institute for Spatial Analysis and Planning in Areas of Intensive Agriculture (ISPA) and goes by the name Virtual GIS (VGIS). VGIS provides an ideal tool box for the environmental modeler, who wants to concentrate on modeling issues rather than the intricacies of GIS.

In the second part, the flow chart-based user interface of VGIS is introduced as a visual programming and prototyping tool similar to STELLA® (HPS, 1994), but working with real GIS data and offering the full functionality of GIS. The conceptual modeling capabilities are exemplified by applications from environmental modeling.

---

1. INTRODUCTION

The domain of universal GIS methods is an area of Geographic Information Science (Goodchild, 1992a) that so far has been neglected by research and for which therefore no body of theory exists. Driven by the heterogeneous market forces every vendor and a multitude of academic developers produced myriads of commands to perform GIS operations. The few existing taxonomies of GIS operations (Aronoff, 1991; Burrough, 1989 and 1992; Goodchild, 1992b; de Man, 1988; Rhind and Green, 1988; Unwin, 1990; Schenkelaars, 1994) are limited either by the data structure that they are based on or by the scope of applications for which they had been developed. They all lack formalization and do not attempt to be truly universal. Section 2 introduces the author's work on a task-oriented systematization of data structure-independent GIS functionality.

On the ecological modeling side, Constanza and Sklar (1985) state that "most ecological modeling work to date has focused on temporal changes. They tend to simulate a point in space and extrapolate the findings for an entire landscape by assuming that the landscape is *homogeneous*. In other words, most models in ecology have little, if any, spatial articulation. It is clear, however, that space needs to be more explicitly included if ecological models are to be truly useful tools for understanding and predicting the behavior of real ecosystems (Risser, *et al.*, 1984)." Although this quote is 10 years old now, it has not lost any of its truth. Section 3 reviews the current state in GIS-based ecological modeling.

Section 4 tries then to apply the universal GIS operations described in section 2 in an environmental modeling context. Special emphasis is given to the phase of conceptual modeling as well as to an object-oriented application of individual modeling units. The conclusion (section 5) describes current limitations to the proto-type and discusses some rather fundamental ideas about the synthesis of GIS and environmental modeling.

2. UNIVERSAL GIS OPERATIONS

One motivation for research on universal GIS operations simplifying that simplify GIS use was the observation, that in spite of its name, current GIS have little to offer to the scientist, who is interested in modeling spatial phenomena. Tomlin's (1990) cartographic modeling language (also known as 'Map Algebra') is the most sophisticated GIS modeling environment so far. This lack has been articulated and mourned by many in the modeling community and resulted in a conference series, devoted to overcome this discrepancy between the GIS and the modeling community (Goodchild *et al*., 1993; Goodchild *et al*., 1995).

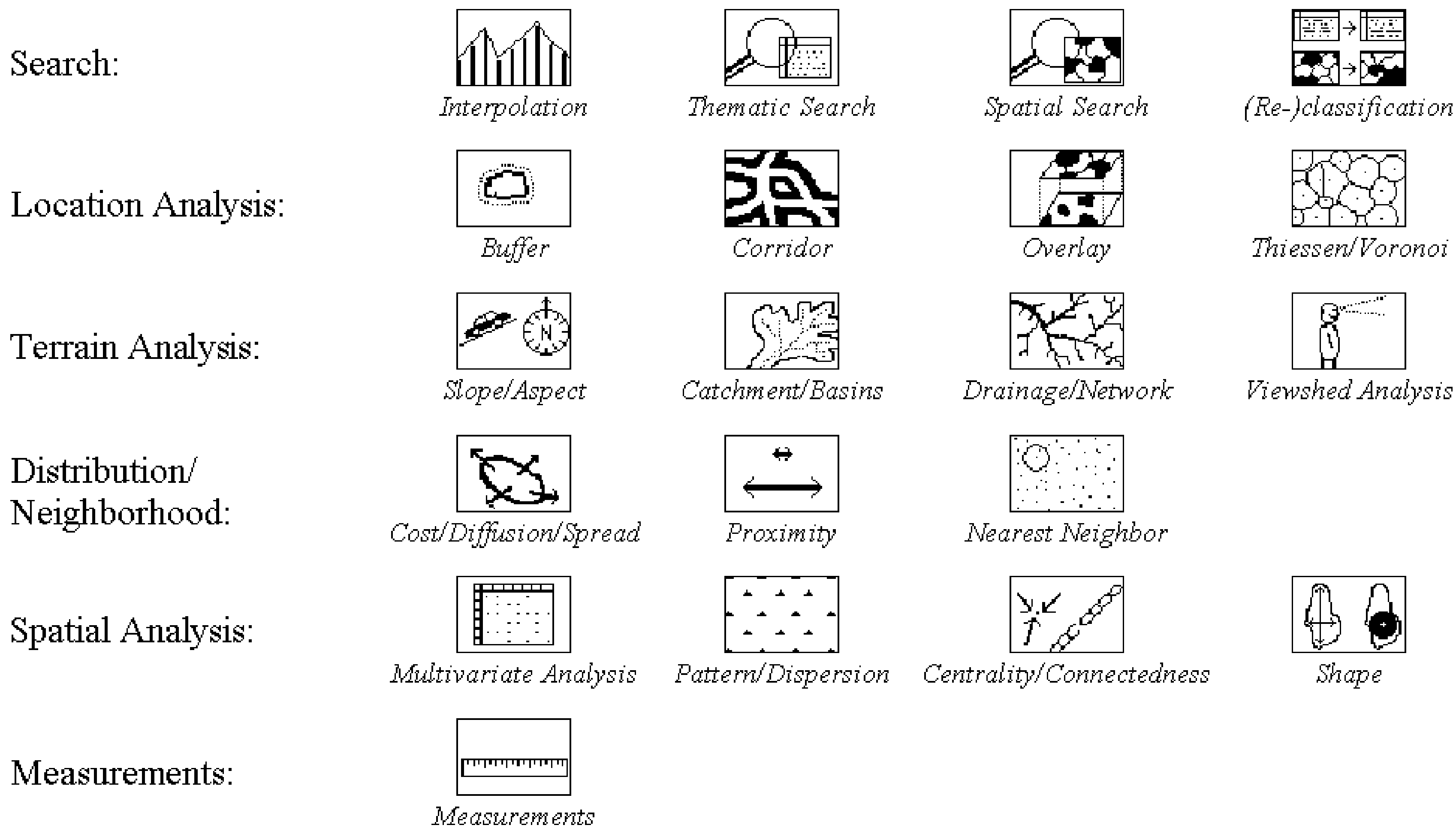
Current Geographic Information Systems (GIS) are so difficult to use that it takes some expertise to handle them, and it is not unusual to assess a whole year until an operator masters a GIS. This is especially cumbersome for cursory users (such as environmental modelers) that employ GIS as one tool among many others. Coulsen, *et al.*, (1991) expressed a similar argument when they wrote "GISs are complex computer programs. Proficient application in natural resource management and landscape ecology involves a commitment to training and practice by the user. None of the GISs would be considered *user friendly* by a human factors engineer." Similar comments may be found throughout (Medyckyj-Scott and Hearnshaw, 1993) and (Turk, 1992).

Although a number of GIS claim to be data structure-independent none of them really is; they all show their origin as so-called either raster (grid cell-based) or vector systems. This data structure distinction has dictated differences in analysis functionality. Even Goodchild (1991a, p. 45) in his often cited classification of spatial data analysis techniques groups them depending on the underlying data model. And while there are numerous efforts to standardize data models (SAIF, SDTS, DIGEST, GDF), so far none of these attempt to standardize the operations as well. The advent of the Open Geodata Interoperability Specification (OGIS) (Buehler, 1994; Buehler, 1995) opens for the first time a real opportunity to develop data structure-independent GIS applications (see also the contribution by Kenn Gardels in this volume). As far as is known so far, however, the Open GIS Consortium (OGC) does not attempt to define high level operations but restricts its specification to low level database (SQL-like) and topological operations based on the work by Egenhofer (1991, 1993).

The author tackled these deficiencies with a user survey to determine *user* expectations of a GIS' functionality (Albrecht, 1995b). The responses reveal a vast array of complexity ranging from elementary operations to compound tasks. A dissection of the latter into fundamental primitives leads towards a normalization of GIS operations. Since the analytic capabilities of GIS are the only ones that distinguishes them from other visualization software or database management systems (Burrough, 1986; Goodchild, 1987), further consideration of GIS functionality within this paper will concentrate on this ability.

By analyzing current GIS user interfaces and omitting all those operations that are due to either the historic development of the particular software package or are a result of the data model employed, the author derived a list of only 20 universal GIS operations that allow to build all but the most exotic GIS applications (see also Table 1). This small set of spatial analytical tools provides the means to perform environmental spatial modeling without having to learn about the intricacies of current GIS. A detailed description of how this particular set of GIS operations was derived is given in (Albrecht, 1996). There, the reader will also find a formalization of these operations based upon a simplified version of the OGIS data model. An implementation of these fundamental GIS operations within an interactive flow-charting environment (see section 4) reveals the window of opportunities opened by this approach.

Table 1. The 20 universal GIS operations



Search:	*Interpolation*	*Thematic Search*	*Spatial Search*	*(Re-)classifi cation*
Location Analysis:	*Buffer*	*Corridor*	*Overlay*	*Thiessen/Voro*

				noi
Terrain Analysis:	*Slope/Aspect*	*Catchment/Basins*	*Drainage/Network*	*Viewshed Analysis*
Distribution/ Neighborhood:	*Cost/Diffusion/ Spread*	*Proximity*	*Nearest Neighbor*	
Spatial Analysis:	*Multivariate Analysis*	*Pattern/Dispersion*	*Centrality/Connectedness*	*Shape*
Measurements:	*Measurements*			

3. THE STATE OF ART IN THE APPLICATION GIS FOR ENVIRONMENTAL MODELING

> *"GIS can be used to seduce the user into an unrealistic sense of model accuracy", using "a few poor, anemic point measurements".*
> (Grayson, *et al.*, 1993, p. 91)

One motivation for the search for the GIS usage simplifying universal GIS operations was the observation, that in spite of its name, current GIS have little to offer to the scientist, who is interested in modeling spatial phenomena. Tomlin's (1990) cartographic modeling language (also known as 'Map Algebra') is the most sophisticated GIS modeling environment so far. This lack has been articulated and mourned by many in the modeling community and resulted in a conference series, devoted to overcome this discrepancy between the GIS and the modeling community (Goodchild, *et al.*, 1993 and Goodchild, *et al.*, 1995).

The two probably most exhaustive overviews (Sklar and Constanza, 1991; Hunsaker, *et al.*, 1993) describe numerous environmental tasks such as inventory, assessment, management, and predicting the fate of environmental resources supporting applications in atmospheric modeling, hydrological modeling, land surface-subsurface modeling, ecological systems modeling, plus integrated environmental models as well as policy considerations for risk/hazard assessment involving these models (see Table 2). All of these, however, use a GIS as an inventory for spatially referenced data and for presentation (map production) only.

Nyerges (1991) identified three primary modes of GIS use, namely *map mode* providing referential and browse information, *query mode* to address specific requests for information based either on field or object views, and *model invocation* which is the only mode that makes use of the analytical capabilities of a GIS. On a rather abstract level Nyerges describes the development of a typical modeling process, however, he fails to note that up to now there exists no GIS that actually supports such a procedure.

Two research projects focus on the problem of facilitating a computational modeling environment employing rather different approaches. The first one to be described here is the computational modeling system (CMS) developed at the Department of Computer Science at the University of California (Smith, *et al.*, 1995). Their computational modeling language (CML) is supposed to support cooperative (geographic) modeling activities at all stages of the modeling process, i.e. data extraction, construction and evaluation of conceptual models, model refinement, and communication of the results (Alonso and Abbadi, 1994). The CML is based upon the concept of so-called representational (or -) structures and their transformations. These -structures contain specifications how to represent the same information using a different data model, so the user does not need to explicitly know about the data model that the source data is based on. Each -structures then also contains the operations that can be applied on it, e.g. a digital elevation model (DEM) may contain the transformations '*union*' to combine several DEMs, '*compute-slope*', or '*max-height*'. Creating such schemas nevertheless requires to learn a new programming language which might be worth the effort for a some model builders, but renders it unlikely that CMS will become a wide-spread tool.

The other research project aimed at facilitating modeling procedures within a GIS is the Virtual GIS (VGIS) project described in Albrecht (1995a). VGIS is a shell that employs the universal GIS operations described in section 2 using a flow charting environment (see Figure 1). Flow charts are a the standard process-oriented tool in visual programming (Chang, 1990; Glinert, 1988; Monmonnier, 1989). As in the case of CMS, the system has yet to prove its usability in real world (meaning non-academic) applications. Proposed (and already granted, see section 6) extensions of this work promise a wider applicability, as one of the interpreters that is intended to be developed for VGIS will be an interface to the OGIS data model.

Table 2. Spatial Models in Different Domains (after Sklar and Constanza, 1991)

Geography

geometric (all descriptive; without feedback loops)

von Thünen (space around a point)

Horton's Law

Christaller (hierarchy of hexagons)

demographic

expansive/relocative

Hägerstrand's innovation diffusion (logistic curve)

Diffusion modeled as a Markovian process using gravity models $T=kW/d^a$ with k and a being empirical measures or spatial stochastic parameters;

i.e. autocorrelation

network (linear programming; assumes linear relationships; no feedbacks)

applications: traveling salesman, commodity flow, spatial allocation, optimization

Hydrology (physical continuity equations)

finite element (hydrodynamic)

link-node design similar to network models in geography

dynamics are a function of the network design and the simultaneous solution of the continuity and momentum equations of every node

no spatial processes in non-link areas, i.e. adjacent to pipelines

solutions to the latter problem:

link size as function of the area surrounding adjacent nodes

parallel pipes

areas adjacent to pipes become reservoirs

finite difference (general circulation)

General Atmospheric Circulation Models (GACM)

3-dimensional mesh; grid along latitude/longitude, 10 or 100 layers

Biology

growth (landscape is an independent exogenic variable)

competitive interaction, i.e. Lotka-Volterra

patch dynamics

interference and exploitation

niche space

population

very similar to geographic demography with more emphasis on density-dependence
in plant ecology: correlation between abundance and spatial heterogeneity;

they emphasize community-level rather than population-level interactions

point-averaged ecosystem

STELLA®-like; condense complex ecosystems into a small number of differential equations

Ecosystem Research (Landscape)

stochastic

fit the definition of <u>spatial</u> modeling (Griffith and Mac Kinnan, 1981)

map the flow of energy, matter and information, designate source, sink and
receptor areas, predict succession in 2D and 3D space, address questions of scale

incorporation of feedback loops, neighborhood influences and spatial im-/exports

results in spatially more articulate models

process-based

analog to the Global Atmospheric Circulation Models above;

grid-type, 2D or 3D flow, material transport as a function of mass balance

only process-based landscape models have the advantage of being spatially explicit, realistic and dynamic

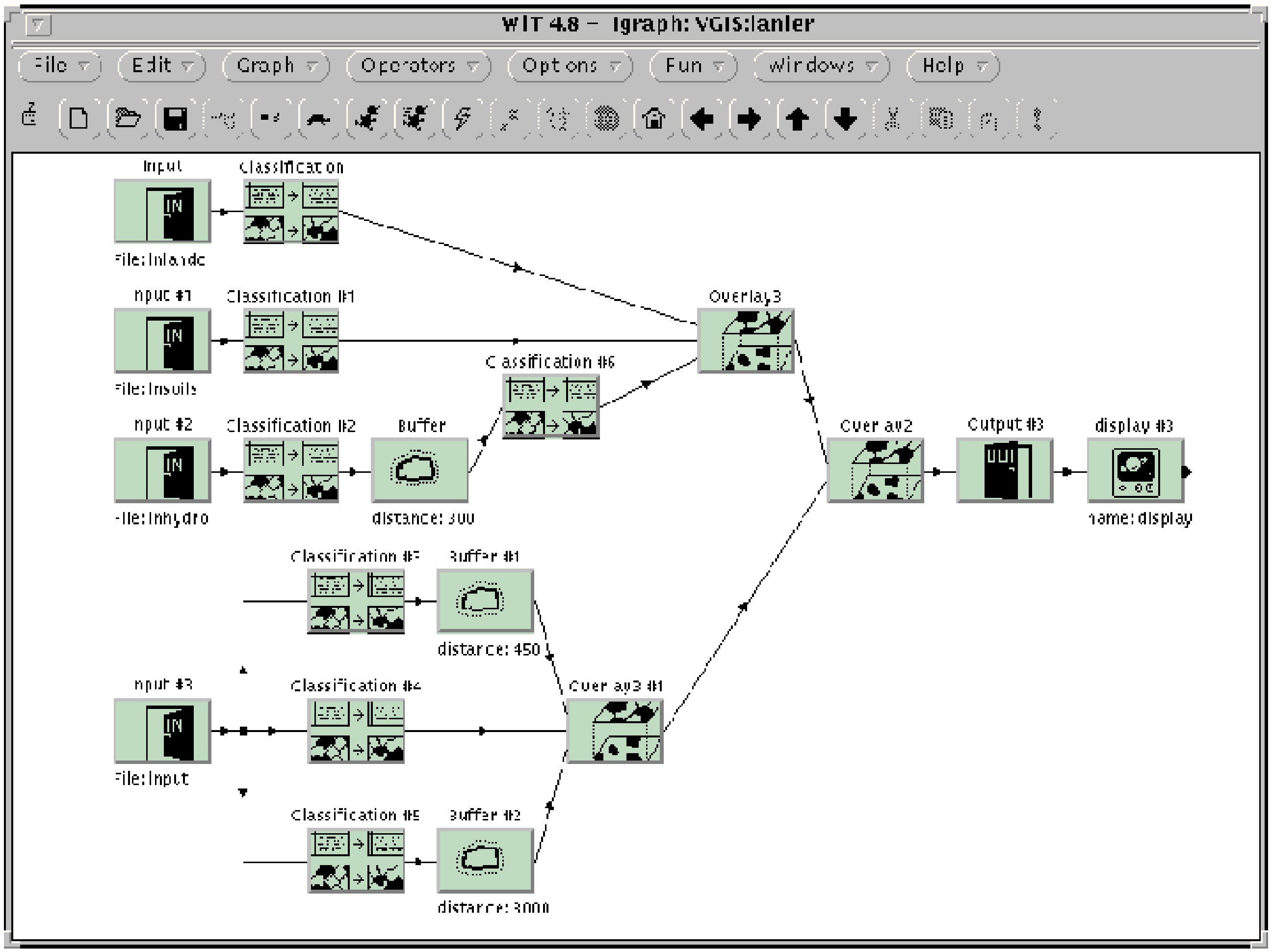


Figure 1. VGIS modeling example (screen dump)

4. ENVIRONMENTAL MODELING WITH VGIS

So far, integrated application of GIS and environmental modeling have been specialized modeling languages, such as PCRaster and DYNAMO in the hydrologic domain (van Deursen and Kwadijk, 1993; Wesseling and v. Deursen, 1995). These languages, however, do not really support the creative process of model building. Rather, they require an intricate knowledge of the model and the language, and are harnessed to fine-tune a fixed model run. The Virtual GIS (VGIS) on the other hand, attempts to be a prototyping tool and development platform similar to STELLA® (HPS, 1994), but working with real GIS data and thereby graphically extending 'Map Algebra' according to the concepts presented in (Kirby and Pazner, 1990). More realistic (real world) data with a locational character have a significant impact on the model results. In addition, geographical displays interactively depicting the nature of the sensitivity of certain parameters can be useful in support of model parameterization. examining scale effects can be accomplished by changing interactively the nature of the data aggregation. The model brings together the locational, temporal, and thematic aspects of phenomena in a geographic process characterization.

Such a visual programming example is depicted in Figure 2, where the modeling flow chart allows the user to "play" with the data flow. Figure 2 represents an intermediate step in the conceptual modeling of erosion. The four input files (rounded boxes) are geology, landcover, precipitation, and elevation. It is possible to model this complex system with only five of the universal GIS operations. Within VGIS, it is easy to test the result of new routing paths within the flow chart. The hypothesis that a certain region buffered around drainage channels has a different water retention capacity can easily be tested by adding one connection to the flow chart. A similar reconfiguration of a conceptual model would require substantial GIS expertise if it were attempted in a vendor GIS. The similarity with the previously mentioned prototyping tool extends to another feature as well: each data object might explode into another model that can either be treated as a black box or specified in a similar manner as its parent level.

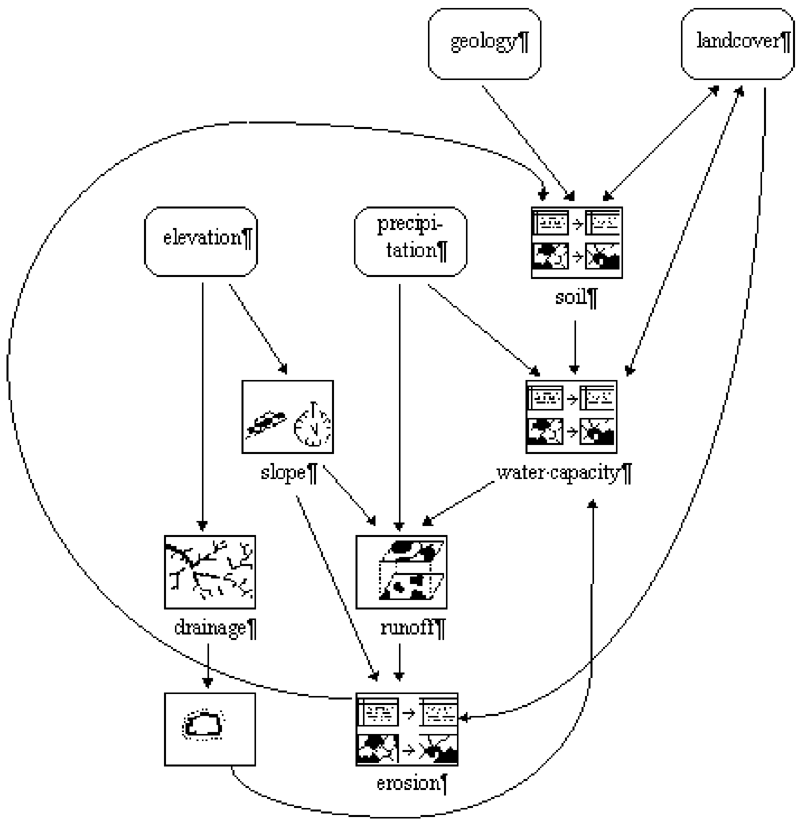


Figure 2. Sample modeling flow chart

Probably the main advantage of the VGIS environment lies in the first time-ever possibility to easily include feedback loops within a GIS. "Landscapes are never static; their elements are in constant temporal and spatial flux" (Merriam, *et al.*, 1991). Sklar and Constanza (1991) therefore consider the incorporation of space as well as time as the most prominent issue in ecosystem research. This needs to be done at all levels of resolution that are meaningful to the myriad ecosystem management problems we now face. It is this explicitly spatial aspect is what motivates landscape ecology.

Another major advantage of the VGIS environment is the inclusion of spatial statistics as GIS functionality. "To consider something a system, it is necessary to describe its boundary and its interaction with the environment" (Frank, *et al.*, 1994). Therefore, one of the first tasks within a modeling environment should be to delineate the spatial boundaries. This can be readily accomplished by calling the '*Pattern/Dispersion*' analysis operation. Although this is one of the core functions of a GIS, the author has yet to come across a reference for an environmental modeling application where a GIS is actually used for this purpose.

5. CONCLUSION

VGIS is implemented as an interpreter to the public domain GIS GRASS. Current work focuses on the development of another interpreter for Arc/Info®. The only handicap for a free distribution of the prototype is the utilization of the flow charting tool WiT®. A detailed technical description is given in (Brösamle, *et al.*, 1996). The VGIS environment does not yet include conditional and iteration operators as they are used in formal programming languages. Therefore, in its current state, the VGIS can not be called a geographic modeling language yet.

One of the main impediments to the application of VGIS in real world applications has been the difficulty to environmental modeling tasks that require the full scope of analytical GIS functionality. Most examples would do well with a minimal set of 'overlay', 'slope calculation' and 'diffusion/spread'. The latter two are often implemented in cell-based modeling systems as well. This experience comes very much as a surprise to a geographer who started out with conviction that space is such an important feature in any environmental model that the full power of GIS will be required to satisfy even minimal scientific deeds. Now the far more modest author recognizes that there is still a long way to go, before we will understand the rules governing landscape ecology before we can implement a modeling system on a truly large scale.

6. ACKNOWLEDGEMENTS

This research is supported by the German Science Foundation which provides the funding for the Virtual GIS (VGIS) project under grant number IIC5-Eh 85/3-2.

7. REFERENCES

Albrecht, J., 1995a. Virtual Geographic Information System (VGIS). In Nunamake, J. and R. Sprague (Eds): *Proceedings 28th Hawaii International Conference on System Sciences*, GIS Minitrack, Volume IV, pp. 141-150. IEEE Computer Society Press, Los Alamitos, CA.

Albrecht, J., 1995b. Semantic Net of Universal Elementary GIS Functions. *Proceedings ACSM/ASPRS Annual Convention and Exposition Technical Papers*, Vol. 4 (Auto-Carto 12), pp. 235-244. Bethesda, MD.

Albrecht, J., 1996. *Universal GIS Operations*. Dissertation, published electronically and accessible as http://www.ncgia.ucsb.edu/~jochen/diss/dissabst.html.

Alonso, G. and A. Abbadi, 1994. Cooperative Modeling in Applied Geographic Research. *International Journal of Intelligent and Cooperative Information Systems*, 3(1):83-102.

Aronoff, S., 1991. *Geographic Information Systems: A Management Perspective*. WDL Publications, Ottawa, 2°, 294 pp.

Brösamle, H., Albrecht, J. and M. Ehlers, 1996. GIS Functionality and Interface Design: The Virtual GIS (VGIS) Example. Invited paper, ISPRS Commission II Working Group II/2, Workshop on New Developments in Geographic Information Systems, Milan, Italy.

Buehler, K., 1994. OGIS Project Document 94-019. *OGIS Geodata Model Overview*. The Open GIS Foundation, Cambridge, MA.

Buehler, K., 1995. Open Geodata Interoperability Specification (OGIS). Presentation given by Kurt Buehler at the OGIS workshop in Zürich, Switzerland, 30 November 1995.

Burrough, P., 1986. *Principles of Geographic Information Systems for Land Resources Assessment.* Monograph on Soil and Resources Survey, No. 12. Clarendon Press, Oxford. 194 pp.

Burrough, P., 1992. Development of Intelligent Geographical Information Systems. *Internatinoal Journal of Geographic Information Systems*, 6(1):1-11.

Chang, S., 1990. *Principles of Visual Programming Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Constanza, Sklar, 1985. Articulation, Accuracy, and Effectiveness of Mathematical Models: a review of freshwater wetlands applications. *Ecological Modeling*, 27(1):45-68.

Coulsen, R., Lovelady, C., Flamm, R., Spradling, S. and M. Saunders, 1991. Intelligent Geographic Information Systems for Natural Resource Management. In Turner, M. and R. Gardner (Eds.): Quantitative Methods in Landscape Ecology, pp. 153-172, New York, Springer-Verlag.

de Man, E., 1988. Establishing a geographic Information System in Relation to its Use. *International Journal of Geographic Information Systems*, 2(3):257.

Egenhofer, M. and R. Franzosa, 1991. Point-Set Topological Spatial Relations. *International Journal of Geographical Information* Systems 5(2):161-174.

Egenhofer, M., Sharma, J. and D. Mark, 1993. A Critical Comparison of the 4-Intersection and the 9-Intersection Models for Spatial Relations: Formal Analysis. Proceedings, *Eleventh International Symposium on Computer-Assisted Cartography* (Auto Carto 11), pp. 1-11.

Frank, A., Egenhofer, M. and D. Hudson, 1994. The Design of Spatial Information Systems. Part 1: Formal Systems. File on ftp://grouse.spatial.maine.edu/pub/SurveyEng/sve451/451.ps.

Glinert, E. (Ed.), 1988. Visual Programming Environments: applications and issues. IEEE Computer Society Press, Los Alamitos.

Goodchild, M., 1991. Spatial Analysis with GIS: problems and prospects. *Proceedings GIS/LIS'91*, pp. 41-48. ASPRS/ACSM, Falls Church, VA.

Goodchild, M., 1992a. Geographical Information Science. *International Journal of Geographic Information Systems*, 6(1):31-46.

Goodchild, M., 1992b. *Spatial Analysis Using GIS*: a seminar workbook, pp. 40-48, 2°, National Center for Geographic Information and Analysis, University of California, Santa Barbara.

Goodchild, M., Parks, B. and L. Steyaert (Eds.), 1993. Environmental Modeling with GIS. Oxford University Press, New York/Oxford.

Goodchild, M., Steyaert, L., Parks, L., Crane, M., Johnston, C., Maidment, D. and S. Glendenning, 1995. GIS and Environmental Modeling: Progress and Research Issues. GIS World, Inc., Ft. Collins, CO.

Grayson, R., Blöschl, G., Barling, R. and I. Moore, 1993. Process, Scale and Constraints to Hydrological Modelling in GIS. In Kovar, K. and H. Nachtnebel (Eds.) *Application of Geographic InformationSystems in Hydrology and Water Resources Management*. Proceedings HydroGIS 93. Vol. 211, pp. 83-92, IAHS Press, Wallingford, England.

High Performance Systems (HPS), 1994. STELLA II: an introduction to systems thinking. Hanover, NH.

Hunsaker, C., Nisbet, R., Lam, D., Browder, J., Baker, W., Turner, M. and D. Botkin, 1993. *Spatial Models of Ecological Systems and Processes: The Role of GIS*. In Goodchild *et al.*, (1993), pp. 248-264.

Kirby, K. and M. Pazner, 1990. Graphic Map Algebra. In Brassel/Kishimoto (Eds.): *Proceedings of the 4th International Symposium on Spatial Data Handling*, Vol. 1, pp. 413-422.

Medyckyj-Scott, D. and H. Hearnshaw, 1993. *Human Factors in Geographic Information Systems*. Belhaven Press, London, 266 pp.

Merriam, G, Henein, K. and K Stuart-Smith, 1991. Landscape Dynamics Models. In Turner, M. and R. Gardner (Eds.): *Quantitative Methods in Landscape Ecology*, pp. 399-416, Springer, New York, NY.

Monmonnier, M., 1989. Graphic Scripts for the Sequenced Visualization of Geographic Data. *Proceedings GIS/LIS'89*, pp. 381-389. ASPRS/ACSM, Falls Church, VA.

Nyerges, T., 1991. Analytical Map Use. *Cartography and Geographic Information Systems*, 18(1):11-22.

Rhind, D. and N. Green, 1988. Design of a Geographical Information System for a Heterogeneous Scientific Community. *International Journal of Geographic Information Systems*, 2(2):175.

Risser, P., Karr, J. and R. Forman, 1984. *Landscape Ecology: Directions and Approaches*. Special Publication No. 2. Illinois Natural History Survey, Champaign, IL (qf. Sklar and Constanza, 1991).

Schenkelaars, V., 1994. Query Classification, a First Step Towards a Graphical Interaction Language. In Molenaar, M. and S. de Hoop, (Eds.): *Advanced Geographic Data Modeling*. Netherlands Geodetic Commission, Vol. 40, pp. 53-65.

Sklar, F. and R. Constanza, 1991. The Development of Dynamic Spatial Models for Landscape Ecology: a review and prognosis. In Turner, M. and R. Gardner (Eds.): *Quantitative Methods in Landscape Ecology*, pp. 239-288, Springer, New York, NY.

Smith, T., Su, J., Abbadi, A., Agrawal, D., Alonso, G. and A. Saran, 1995. Computational Modeling Systems. *Information Systems*, 20(2):127-153.

Tomlin, D., 1990. *Geographic Information Systems and Cartographic Modeling*. Prentice-Hall, Englewood Cliffs, NJ.

Turk, A., 1992. *GIS Cogency: cognitive ergonomics in geographic information systems*. Unpublished dissertation, Department of Surveying and Land Information, University of Melbourne.

Unwin, D., 1990. A Syllabus for Teaching Geographical Information Systems. *International Journal of Geographic Information Systems*, 4(4):461-462.

van Deursen, W./Kwadijk, J., 1993. RHINEFLOW: an Integrated GIS Water Balance Model for the River Rhine. In *Applications for Geographic Information Systems in Hydrology and Water resources*. IAHS Publications No. 211, pp. 507-518.

Wesseling C., and W. van Deursen, 1995. A Spatial Modelling Language for Integrating Dynamic Environmental Simulations in GIS. *Proceedings Joint European Conference and Exhibition on Geographical Information*, Vol. 1, pp. 368-373.

Jochen H. Albrecht
Institute for Structural Analysis and Planning in Areas of Intensive Agriculture (ISPA)
University of Vechta
Postfach 1553
49364 Vechta, Germany
Telephone: +49 (444) 11 53 37
FAX: +49 (444) 11 54 45
http://www.geog.ucsb.edu/~jochen/

Next Up Previous

Next: Environmental Impact Assessment **Up:** Spatial Reasoning for Environmental
Previous: Spatial Reasoning for Environmental

Introduction

Geographical Information Systems (GIS) in which geo-coded data sets are manipulated and displayed are a powerful means of managing and analyzing spatial information, and are widely used in environmental planning and monitoring to aid planners to make deci sions and predictions. They are used increasingly by those involved in the preparation and interpretation of Environmental Impact Assessment reports. More effective support would be provided by a computer system which is able to capture and realize the reasoning processes involved in the planning process as well as storing and keeping track of the dynamic spatial data.

A severe limitation of Geographic Information Systems is their lack of support for reasoning about the spatial entities which they represent. Knowledge-Based systems (KBS) are well established and the technology well understood, but there has been little investigation of their amalgamation with GIS, although most modern GIS incorporate analytical models with an interface to powerful query langu ages (Kowalski 1986) and there is a large body of research into topological spatial relations such as (Clementini 1994) The lack of investigation into KBS/GIS integration is because, hitherto, the data representation techniques used in GIS have not been directly compatible with the representations used by existing approaches to spatio-temporal reasoning (Hayes, 1979, Allen, 1983, Kowalski 1986). The integration of GIS with knowledge-based systems would provide a means of supporting experts making decisions and predictions in the domain of dynamic spatial data.

The two types of data which are in common use in GIS are vector data, which originates from cartographic maps and point-referenced spatial data sets, and raster data from the growing number of remote sensors. A technique for the representation of data onto which raster representations map naturally is that of *tesserals* in which space is divided hierarchically into tiles (Morton 1966, Grunbaum, 1987). It has been suggeste d by (Diaz 1986) and (Bell, 1983) that it could be used as a data representation for GIS.

SPARTA (SPAtial Reasoning with Tesseral Addressing) is a tesseral spatial reasoning system which incorporates a technique, based on the quad tesseral representation of space, which operates a constraint propagation mechanism.

The data representation enables two dimensional (and higher) space to be treated linearly, allowing the reasoning technique to operate using only the constraints *isWithin*, *isSubset* and *isEqual* (and their negations). The technique has been used successfully to resolve two-dimensional reasoning scenarios and its use is illustrated in this paper by its application to the search for suitable building development sites which fall within acceptable noise pollution levels generated by traffi c from a nearby major road.

Next Up Previous

Next: Environmental Impact Assessment **Up:** Spatial Reasoning for Environmental
Previous: Spatial Reasoning for Environmental

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next | Up | Previous

Next: Data Representation - **Up:** Spatial Reasoning for Environmental **Previous:** Introduction

Environmental Impact Assessment Reports

It is a generally accepted principle that our environment must be protected in order to preserve human health and quality of life, and to maintain the diversity of species and the reproductive capacity of the ecosystem as a basic resource for life. Large development projects may have adverse effects on the environment, and the relevant authorities make judgements about such effects when scrutinizing the proposals before granting permission to proceed. The developer is therefore required to submit, with the planning application, a statement of an assessment of the environmental impact of the development proposed, together with a description of mitigating measures.

The types of project and the environmental factors which they affect are many and varied. One of the most frequent categories of development for which EIARs are produced is the construction of roads and motorways, in which it is mandatory to consider the effects of a major development on ambient noise levels. We may take as a fairly typical example, the planning of a new motorway, together with the development of the area through which it passes. The developers need to be able to identify alternative routes for the road which do not substantially affect existing development in terms of the amount of noise generated, and they need to determine the appropriate areas for new housing development which fall outside the envelope of unacceptable noise level due to the road.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Next: Quad Tesseral Addressing **Up:** Spatial Reasoning for Environmental **Previous:** Environmental Impact Assessment

Data Representation - Quad Tesserals

In order to be able to reason directly about geographic data in a GIS, its representation must be compatible with that required by the reasoning system itself. The Quad Tesseral Addressing technique which provides this compatibility is briefly described here, and interested readers are directed to (Diaz 1986) for further information.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Next: Linearization by Morton **Up:** Data Representation - **Previous:** Data Representation -

Quad Tesseral Addressing

Quad tesseral addressing is an alternative method of defining points in two-dimensional Euclidean space which has some significant advantages over the Cartesian approach. An important feature of Tesseral addressing is the ability to map every part of the spatial domain onto a one-dimensional sequence, and when stored with attributes in a database, each address can act as a single key to data. Tesseral addressing is hierarchical, the place system being used to indicate hierarchical depth, and each full tesseral address associated with a tile is unique. Furthermore, tesseral addressing can be extended to three or four dimensions using additional symbols (Diaz 1986, Grunbaum, 1987), and there are *arithmetics* which support tesseral addressing. for example, addition is tile translation.

Tesseral addresses are generated by quartering the positive quadrant of Cartesian space to yield *parent* tiles labeled as shown in Figure 1, which illustrates one explanation of quad tesseral addressing. We can continue this process of hierarchical division, with new tesseral addresses generated by always appending to the right of the parent tesseral address, until some required depth is reached.

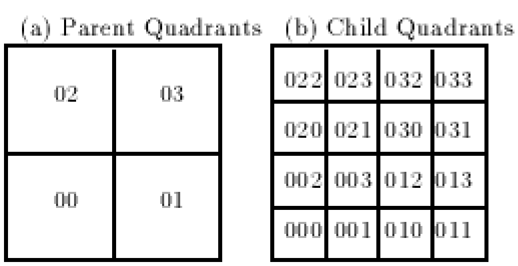


Figure 1: Basic Tesseral addressing

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Next: Reasoning with Tesserals **Up:** Data Representation - **Previous:** Quad Tesseral Addressing

Linearization by Morton Sequencing

In the positive quadrant, the numbering of tesseral addresses follows an ordering known as the Morton sequence (Morton 1966) illustrated below in Figure 2. The *Morton number* of a tile is its index in this sequence, and is related to the tesseral address at the level of its binary representation (Diaz 1986).

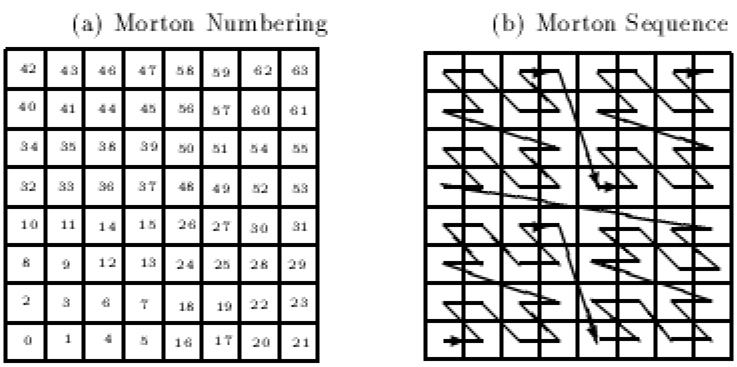


Figure 2: Morton Numbering

The importance of the Morton sequence is that we are able to linearize space - imagine space to be a packed ribbon which can be unraveled - and this in turn enables us to apply one-dimensional reasoning techniques. Moreover, the linearization extends to three, four and higher dimensions, applying the same technique and avoiding the combinatorial explosion of relations (see section 4).

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Next: Overview **Up:** Spatial Reasoning for Environmental **Previous:** Linearization by Morton

Reasoning with Tesserals

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Overview

In the SPARTA system, objects of interest, such as a lake or an area which contains housing, are represented by a name and a set of tesseral addresses, in Morton sequence, which describes where they are *or where it is possible for them to be*.

Spatial problems are presented to the SPARTA system in the form of scripts comprising a set of object declarations and a set of *constraints*. For example, given an area such as the Wirral peninsula in Merseyside, bounded partly by the rivers Mersey and Dee and the Irish Sea, and comprised of urban and rural areas, it can be represented as shown in Figure 3 as a set of objects such as:

object water [location]

object urban [location]

object rural [location]

object M53 [location]

object NewRoad [shape] [location]

where the *shape* and *location* of the object are comprised of sets of tesseral addresses.

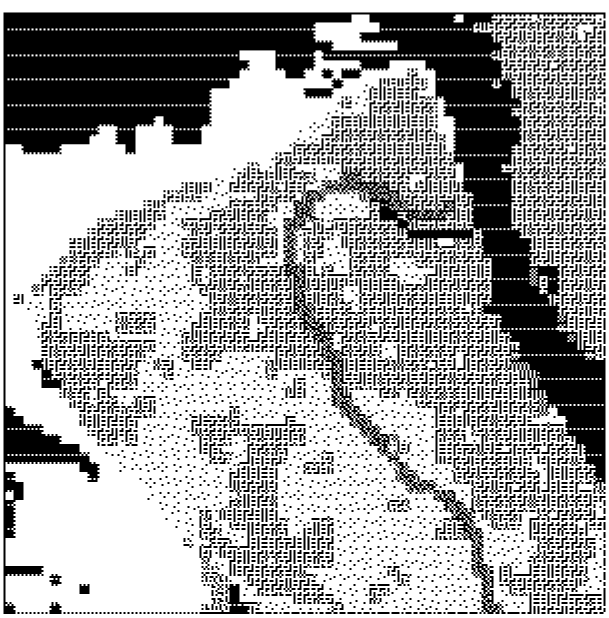


Figure 3: Tesseral Map of The Wirral

The *shape* of the object (omitted where the location of an object is known) may be presented to the system as one of the following constructs, or as a combination of one or more:

- A single tesseral address, which represents one tile.
- A sequence of tesseral addresses defined by a start and an end address, for example: [010..021] which includes addresses 010,011,012,013,020 and 021.
- A *box* of tesseral addresses, defined by the bottom left and top right corner addresses (box 001 012).
- A *line* of addresses defined by its two end addresses (line 0011 0312).
- A *circle* of addresses, defined by a tile whose centroid is the centre of the circle, and an address which represents the radius, measured by its distance from the tile zero. (circle 0312 0010 represents a circle centred on tile 0312 with a rad ius of 2 tiles.)

The shape of an object is defined with reference to the origin of the object space, which is the zero tile, so that the minimum bounding box of the shape is placed with its bottom left corner at zero. The *location* of the object is its actual position in the space, if known absolutely, or its possible locations where more than one possibility exists, in the form of Morton-ordered lists. Such a linear storage structure facilitates simple comparison of locations.

Next Up Previous

Next: Types of Objects **Up:** Reasoning with Tesserals **Previous:** Reasoning with Tesserals

---

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Next: Constraints **Up:** Reasoning with Tesserals **Previous:** Overview

Types of Objects

Two distinct types of object may exist:

- **Tile objects**: The shape is defined by a single tesseral address.
- **Zone objects**: The shape is comprised of more than one address.

and each of these types of object may be categorized as:

- **Fixed** : A tile or zone object with known location.
- **Free** : A tile or zone object which has more than one possible location.

Then the declaration:

object: A [0] [[000..003], [021]]

means that A is a tile object, which may be at any of the addresses 000, 001, 002, 003 or 021.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Constraints

The description of a spatial problem includes a section in which the known relationships existing between the objects are expressed. The system recognizes six types of constraint: isWithin, isnotWithin, isSubset, isnotSubset, isEquals and isnotEquals. ``i sWithin'' and ``isnotWithin'' are functions which deliver alternative sets of addresses. Each of the remaining four constraints is interpreted as a Boolean constraint, so that a question such as ``Is area1 equal to area2'' can be expressed as *constraint: area1 equals area2*.

Operands for constraints are comprised of sets of tesseral addresses. In order to be able to express all possible relations which exist in 2-dimensional space, there are two mechanisms for defining sets of addresses:

- **Object Name** The description of an object's location, possible or fixed, allows the expression of simple relations such as *during* and *contains*. Thus given the object declarations :

  object: A [002]

  object: B [033]

  object: C [0] [000..003]

  Object A occupies a tile at address 002, object B at 033 and C occupies a tile which may be at any of the tiles 000..003.

  The constraint: *constraint: A isSubset C* would translate to: 002 $\subseteq$ 000..003 so the predicate would return ``true''.

- **Object Name and Offset** Areas may be defined in terms of an *offset* applied to an object's set of location tesseral addresses. More complex relations, such as *B northeast of A* are possible with this mechanism. Then:

  constraint: B isSubset A offsetBy [box 003 031]

  states that an offset, [box 003 031] is first to be applied to the addresses of the location associated with object A before implementing the test for *isSubset*, i.e. adding (tesserally) the addresses of the box, which are 003, 012..013, 021, 030.. 031 to the location of A, which is 002, results in the set 021, 023..033. The constraint is then interpreted as: 033$\subseteq$021, 023..033

which in this case returns ``true''.

Next Up Previous

Next: Constraint Satisfaction **Up:** Reasoning with Tesserals **Previous:** Types of Objects

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Next: Operation **Up:** Reasoning with Tesserals **Previous:** Constraints

Constraint Satisfaction

Spatial problem scenarios are presented to the system in the form of groups of constraints which the system will attempt to satisfy. The system first determines the arguments for the constraints, as described in section Constraints, and treats the relations as predicates, i.e. they either succeed or fail. Where a predicate succeeds, the locations associated with its two arguments represent a solution to the constraint and are stored in a *solution space*, *S*, represented as a predicate of the form: *configuration(S)* where *S* comprises a lis t, stored as a tree, of the objects under consideration together with their candidate sets of tesseral addresses. Initially this list will be empty. As each constraint is considered in turn this will result in information being added to, or updated in *S*. This process will continue until all constraints have been considered and all possible solutions for each constraint have been generated. The solution in *S* will represent a final solution.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Operation

The system's operation is illustrated by the following simple scenario. The system has arrived at the point where object A has three possible locations associated with it and B has four.

object: A [[000,001],[002,003],[020,021]]

object: B [[000..003],[010..013],[020..023],[030..033]]

Then given the constraint: *constraint: A isSubset B* each possible location of A must be tested against each location for B:

Only one location for B is compatible with each location of A. By grouping the results we arrive at two possible solutions:

object: A [[000,001], [002,003]]

object: B [000..003]

object: A [020,021]

object: B [020..023]

This result should be interpreted as follows: If A is at locations 000, 001 or 002,003, then B can only be at 000..003. If A is at 020, 021, B can only be at 020..23. Thus we are presented with alternative solutions to the problem presented to the system in terms of the possible locations for A and B, related by the given constraint.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Next: Calculation of Noise **Up:** Spatial Reasoning for Environmental **Previous:** Operation

SPARTA and Noise Levels

We apply the SPARTA system to the solution of the types of problems which arise due to the increased noise level within the environment of a major road. Using SPARTA, we are able to identify candidate areas for development near a major road. This involves finding those areas for which the noise generated by the traffic flow falls below a certain specified level to maintain the ambient noise level within a cceptable limits. We are then able to calculate which areas previously unsuitable are brought into the acceptable domain by the construction of a barrier which minimizes the effects of noise. For projects in which the noise levels fall outside the acceptable limits, this enables developers to specify the mitigation measures to be applied in order to bring the levels inside the limits.

We examine two scenarios :

- The developers wish to identify areas near an existing or proposed new road which, if developed for housing, school, hospital etc., would fall within the acceptable noise level bands.
- Given a set of criteria, the developers wish to find the most acceptable route to build a new road.

Note that in current practice, the route or development site is chosen and the effects of noise determined, rather than the other way round, in which the areas which fall within acceptable levels are identified as candidates for development.

We next examine the method of calculating acoustic levels due to road traffic, then show how this is incorporated into the SPARTA system to enable it to deliver solutions to the problems described above.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Next: Using SPARTA to **Up:** SPARTA and Noise **Previous:** SPARTA and Noise

Calculation of Noise Levels due to Major Road

In the United Kingdom, procedures are laid down by the Department of Transport (DoT) (DoT) in order to obtain the level of noise generated by a road, calculated at a particular reception point. The procedures exist primarily for the determination of entitlement to compensation under the Noise Insulation Regulations. However, they are also intended to provide guidance for calculating traffic noise for more general applications such as planning fo r land use, highway design and, significantly for this work, environmental appraisal of road schemes.

The noise levels are categorised according to the decibel measure (dB(A)) where, in general, level A at less than 55 dB(A) represents the acceptable limits for noise, with no compensation liable for mitigation, and for which planning permission would prob ably not be refused on grounds of noise, with Level D at over 72 dB(A) definitely above the limits for approval.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Using SPARTA to find Noise Level Envelopes

When presented with a spatial problem, we have seen that the SPARTA system delivers solutions as sets of alternative addresses, generated from a number of declarations of objects and constraints between them. These addresses are easily presented to the us er as a map.

The integration of functions such as that of noise level against distance from the road source enables the reasoning technique to be applied to data stored in the tesseral GIS. From this it follows that we are able to reason about areas related by their acoustic level, and obtain answers to questions such as "where are the sites suitable for housing development if a major road is constructed through this area?"

We can also reason about the effects of mitigation of noise levels by erection of barriers at suitable sites - ``How high and where should a barrier be to bring this area into noise level A'' and ``With this barrier at this site, where are the alternative sites for a hospital'. This is accomplished by solving the appropriate acoustic equation with corrections applied for the barrier information.

We illustrate this functionality here. We define a problem in which objects are declared to represent a road, which generates noise, areas which are suitable for building, and areas which have other features which are relevant to the problem. The area which is sought for a development site must be within 5 km of the road. We then apply constraints which are couched in terms which define the restrictions we are seeking, and which yield all the addresses which fulfil the constraint.

The level of noise at a particular point along a road is a function of its distance from the source of the noise. Solving the equation for a given noise level delivers the required distance from the source at which this level is found. This distance is then used to calculate the limits for the various levels of noise, where development would not be allowed to proceed.

For example, assuming no unusual circumstances (such as limited angle of view or steep gradient), traffic flow of 1700 vehicles/ hour at a speed of 91km/h, 10 percent of which are heavy vehicles, then the distance from the noise source at which category A noise level of less than 55 dB(A) applies is calculated to be 673 metres. This distance is converted to the appropriate number of tesseral tiles according to the scale, for use as the radius of a circle as described in the overview to this section, and forms part of a constraint.

The tesseral map of the Wirral peninsular as shown in Figure 3 is declared in the SPARTA system as:

object: noiseA

object: water *location*

object: rural *location*

object: urban *location*

object: M53 *location*

object: forDevelopment

and the constraints:

constraint: forDevelopment notWithin urban

constraint: forDevelopment notWithin M53 offsetBy circle radius1

where radius1 is the distance calculated as above.

constraint: forDevelopment within M53 circle radius2

where radius2 is 5km converted to tesserals.

Solutions after applying the first constraint are the tesseral addresses of all the area which is not greenbelt land, and this set of addresses is now associated with the object forDevelopment. Applying the second constraint further prunes the set to include only those which fall outside the acceptable noise level generated by the road, and the third constraint delivers the final solution of all those addresses which are within 5 kilometres of t he road. These addresses are displayed as a map, thus providing the user with visual information which supports the planning process.

With the incorporation of functions to calculate the effects of introducing mitigating measures for the acoustic effects of the road, such as the erection of barriers close to it, the system is able to present the user with an alternative map of suitable areas. For example, with the erection of a 3.5 m high noise insulation barrier situated 5.0 metres from the edge of the carriageway, the distance at which noise level category A occurs is reduced to 55 metres. The new distance is input and the new map displayed.

Inputting values of the decibel levels for the other noise categories enables the users to produce a map of the noise levels around the road.

Next Up Previous

Next: Conclusions and Further **Up:** SPARTA and Noise **Previous:** Calculation of Noise

---

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Conclusions and Further Work

This paper has described a spatial reasoning system in which two-dimensional spatial data is represented as Quad Tesseral addresses. An important advantage of this data representation is the linearisation of space in a way which enables one-dimensional reasoning techniques to be applied, thus avoiding the combinatorial explosion of relations which are necessary in higher dimensions. The simple manipulation of groups of Quad Tesseral addresses is possible due to the associated Tesseral arithmetic and Morton sequencing.

This representation enables the user to reason directly about spatial - and therefore geographic - data, such as that manipulated in a GIS, and will facilitate and support the decision and prediction process. Thus it is envisaged that this integration of GIS and spatial reasoning will lead to the implementation of a system which will aid planners and developers who are involved in the preparation and interpretation of Environmental Impact Assessment Reports.

ACKNOWLEDGEMENTS

The work described here is being carried as part of the d-GKBIS project and is being supported by the Engineering and Physical Science Research Council.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

References

Allen, J.F. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11), November 1983.

Bell, S.B.M., Diaz, B.M., Holroyd, F. and Jackson, M.J. Spatially referenced methods of processing raster and vector data. *Image and Vision Computing*, 1:211-220, 1983.

Department of Transport. *Calculation of road traffic noise.* HMSO, Welsh Office.

Diaz, B.M. and Bell, S.B.M. *Spatial Data Processing using Tesseral Methods*. NERC Unit for Thematic Information Systems, Swindon, UK., 1986.

Clementini, E, Sharma, J and Egenhofer, M.J. Modelling Topological Spatial Relations: Strategies for Query Processing *Comput. & Graphics*, 18(6), 1994.

Grunbaum, Branko and Shephard, G.C. *Tilings and Patterns*. Freeman, 1987.

Hayes, P.J. The naive physics manifesto. In D. Mitchie, editor, *Expert Systems*. Edinburgh University Press, 1979.

Kowalski, R. and Sergot, M. A logic-based calculus of events. *New Generation Computing*, 4, 1986.

Morton, G.M. A computer oriented geodetic data base, and a new technique in file structuring. *IBM Canada Ltd.*, 1966.

B.Beattie, F.Coenen, T.J.M.Bench-Capon, B.M.Diaz and M.J.R.Shave.
Department of Computer Science, University of Liverpool,
P.O.Box 147, Liverpool L69 3BX U.K.
email: bbeattie@csc.liv.ac.uk
tel: 0151 794 3698 fax: 051 794 3715

A.Hough.
SGS Environment Ltd.
Yorkshire House, Chapel St. Liverpool, U.K.

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995

Next Up Previous

Up: Spatial Reasoning for Environmental **Previous:** References

About this document ...

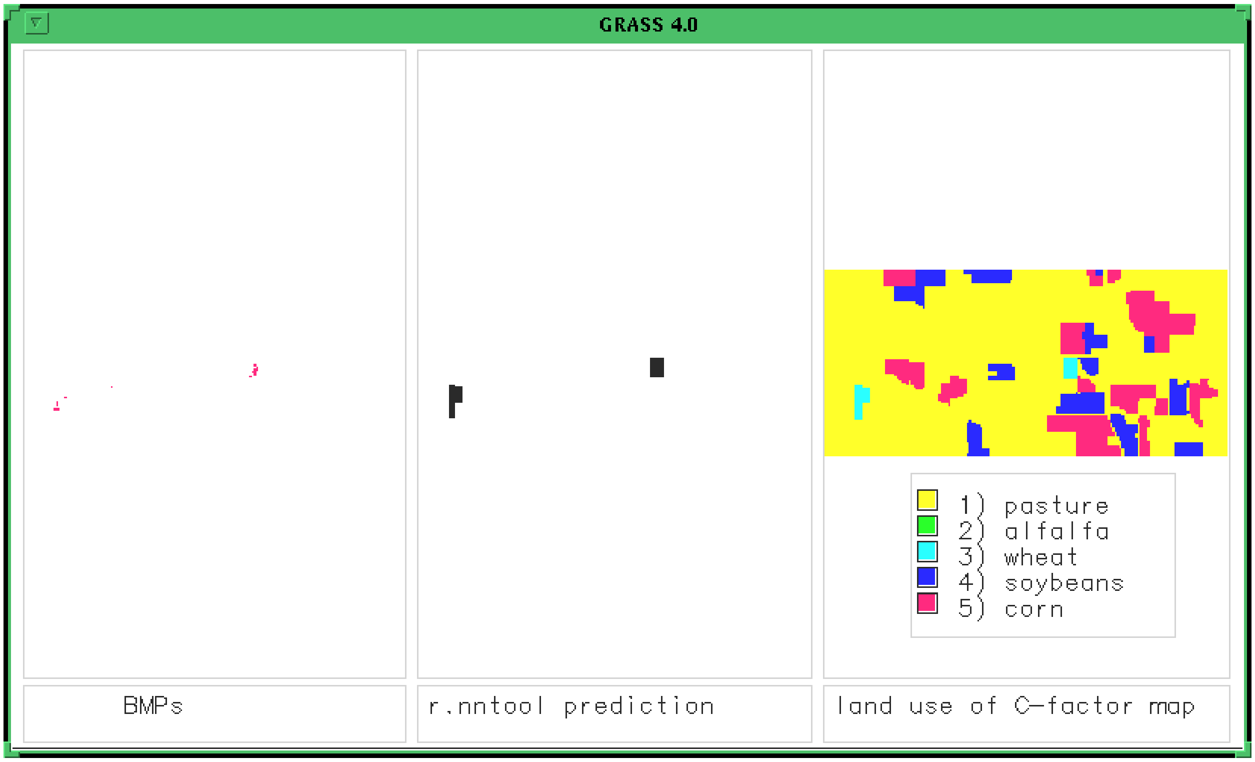
Spatial Reasoning for Environmental Impact Assessment

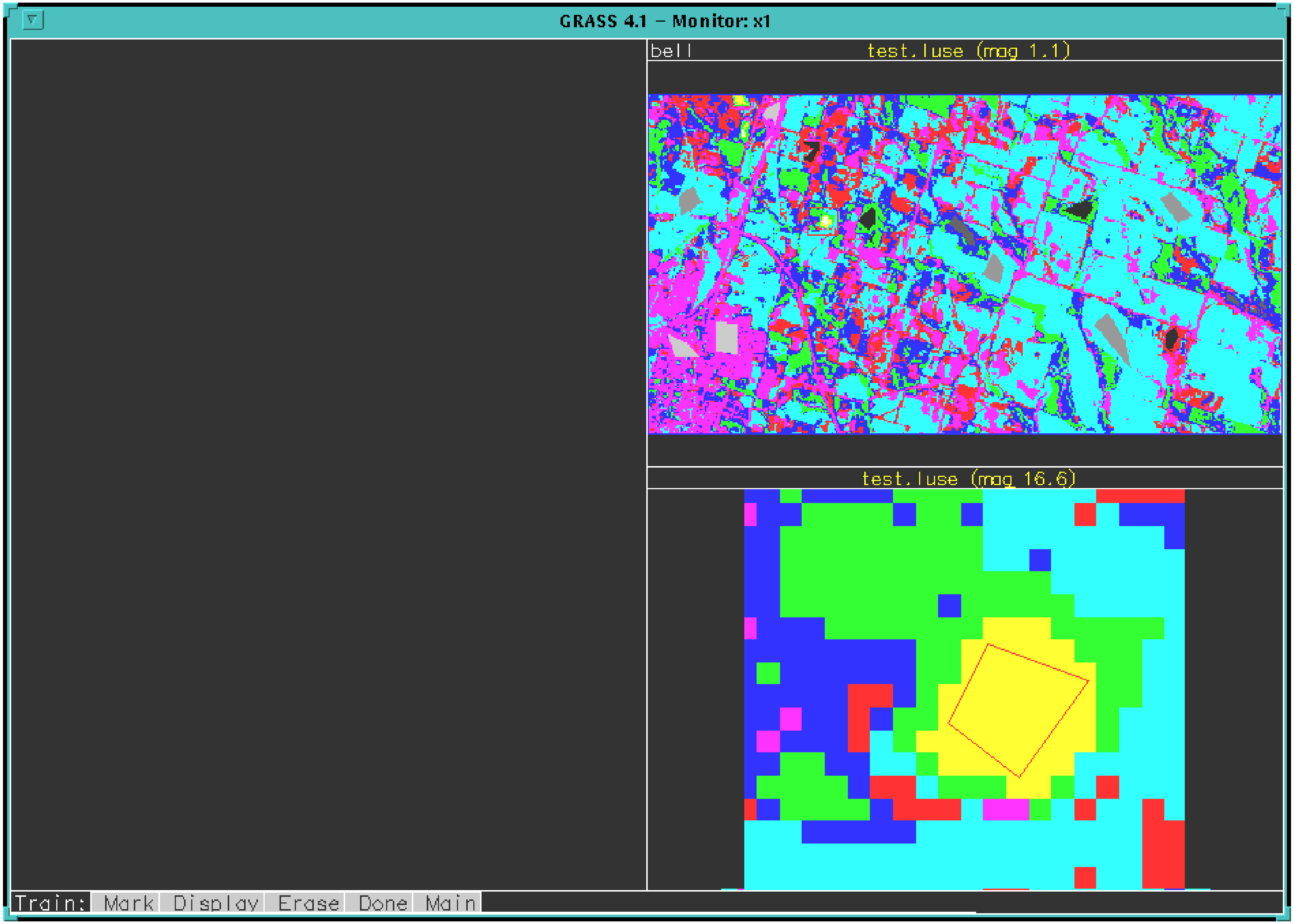
This document was generated using the **LaTeX**2HTML translator Version 0.6.4 (Tues Aug 30 1994) Copyright © 1993, 1994, Nikos Drakos, Computer Based Learning Unit, University of Leeds.

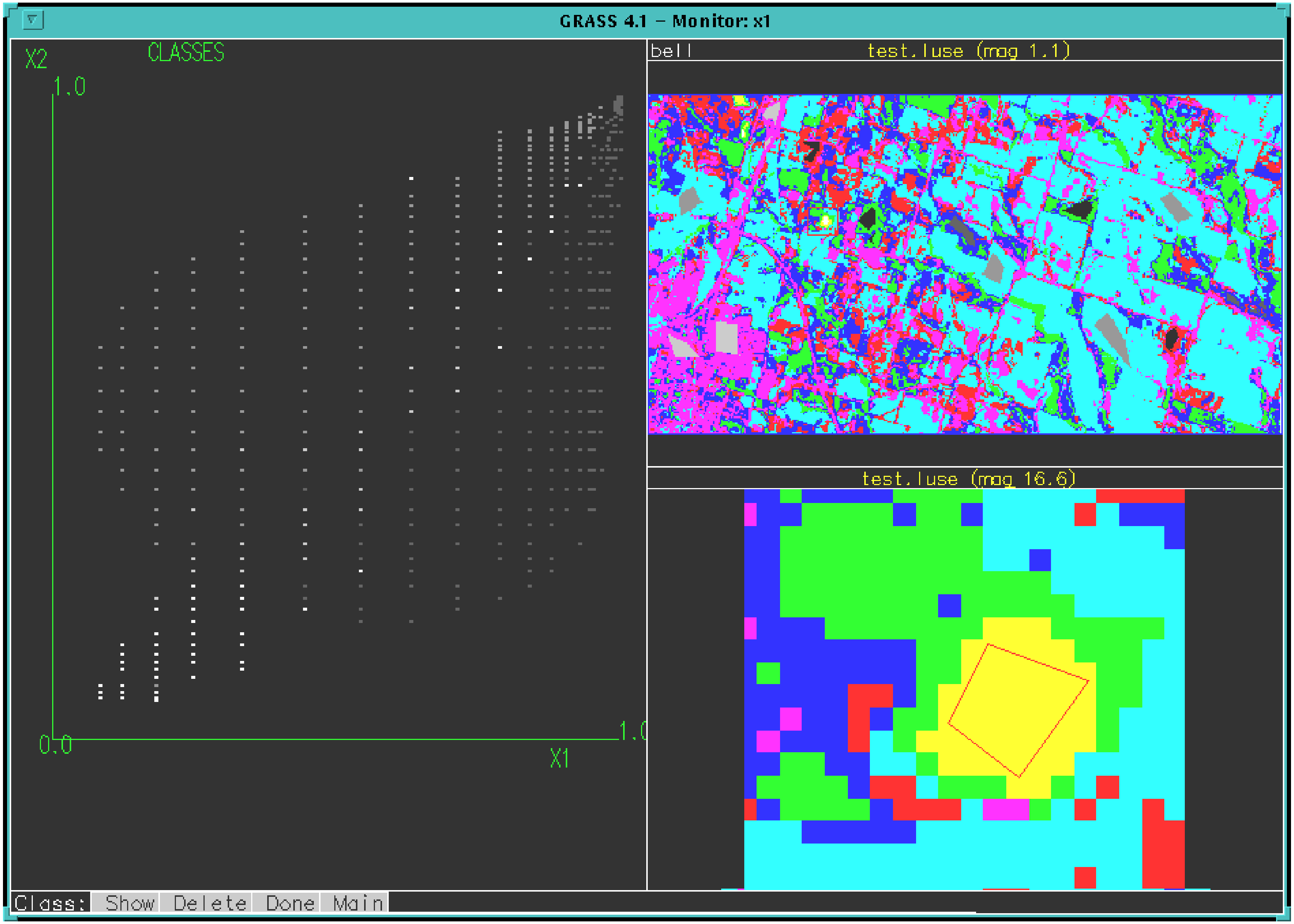
The command line arguments were:
latex2html `santapap.tex`.

The translation was initiated by B.J.H. Beattie on Tue Nov 21 17:11:01 GMT 1995

B.J.H. Beattie
Tue Nov 21 17:11:01 GMT 1995


GRASS 4.0
1) pasture
2) alfalfa
3) wheat
4) soybeans
5) corn
BMPs
r.nntool prediction
land use of C-factor map


GRASS 4.1 – Monitor: x1
bell
test.luse (mag 1.1)
test.luse (mag 16.6)
Train: Mark Display Erase Done Main


GRASS 4.1 – Monitor: x1
CLASSES
X2
1.0
0.0
X1
1.0
bell
test.luse (mag 1.1)
test.luse (mag 16.6)
Class:
Show
Delete
Done
Main

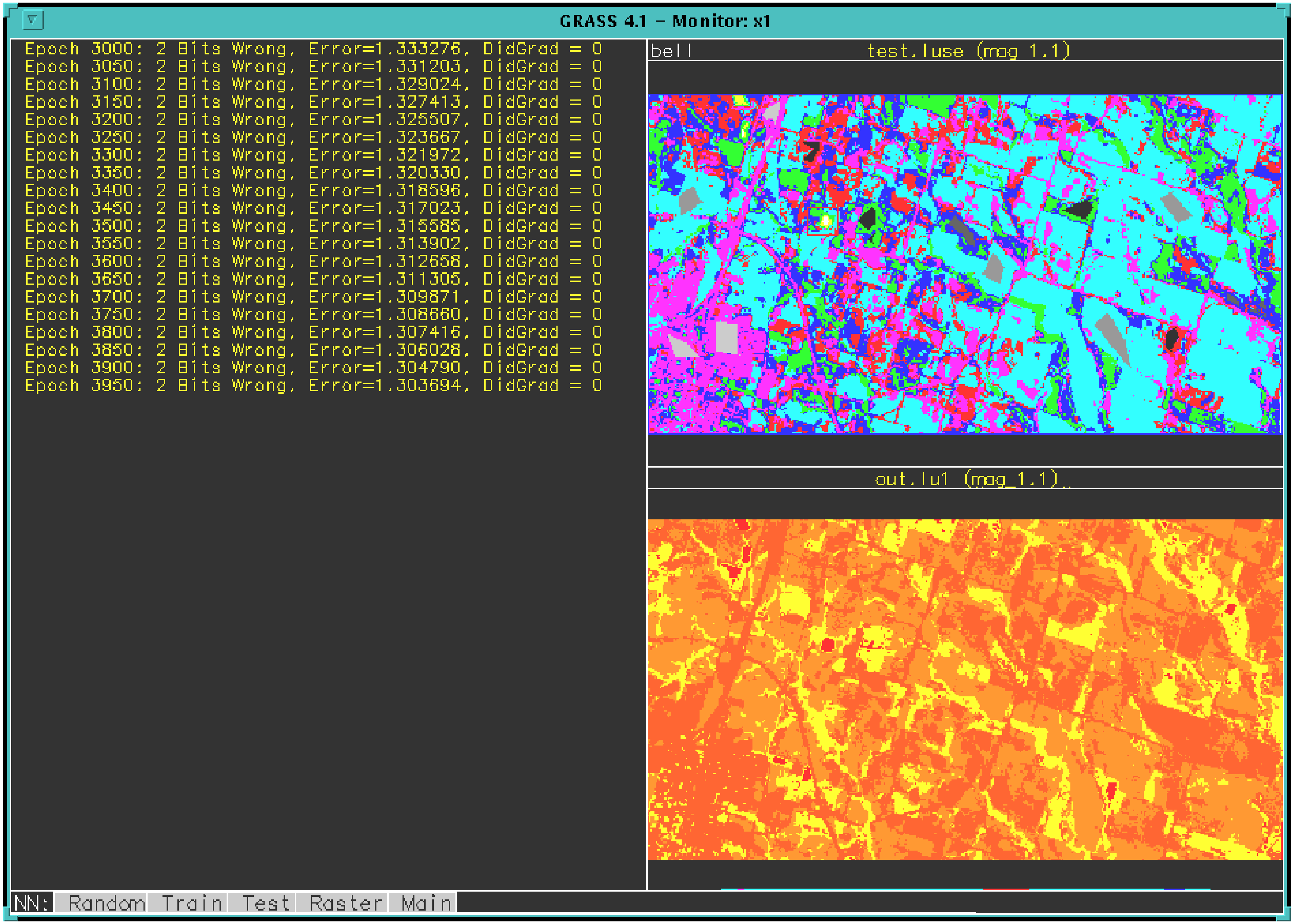

GRASS 4.1 – Monitor: x1
Epoch 3000: 2 Bits Wrong, Error=1.333276, DidGrad = 0
Epoch 3050: 2 Bits Wrong, Error=1.331203, DidGrad = 0
Epoch 3100: 2 Bits Wrong, Error=1.329024, DidGrad = 0
Epoch 3150: 2 Bits Wrong, Error=1.327413, DidGrad = 0
Epoch 3200: 2 Bits Wrong, Error=1.325507, DidGrad = 0
Epoch 3250: 2 Bits Wrong, Error=1.323667, DidGrad = 0
Epoch 3300: 2 Bits Wrong, Error=1.321972, DidGrad = 0
Epoch 3350: 2 Bits Wrong, Error=1.320330, DidGrad = 0
Epoch 3400: 2 Bits Wrong, Error=1.318596, DidGrad = 0
Epoch 3450: 2 Bits Wrong, Error=1.317023, DidGrad = 0
Epoch 3500: 2 Bits Wrong, Error=1.315585, DidGrad = 0
Epoch 3550: 2 Bits Wrong, Error=1.313902, DidGrad = 0
Epoch 3600: 2 Bits Wrong, Error=1.312658, DidGrad = 0
Epoch 3650: 2 Bits Wrong, Error=1.311305, DidGrad = 0
Epoch 3700: 2 Bits Wrong, Error=1.309871, DidGrad = 0
Epoch 3750: 2 Bits Wrong, Error=1.308660, DidGrad = 0
Epoch 3800: 2 Bits Wrong, Error=1.307416, DidGrad = 0
Epoch 3850: 2 Bits Wrong, Error=1.306028, DidGrad = 0
Epoch 3900: 2 Bits Wrong, Error=1.304790, DidGrad = 0
Epoch 3950: 2 Bits Wrong, Error=1.303694, DidGrad = 0
bell
test.luse (mag 1.1)
out.lu1 (mag_1.1)
NN: Random Train Test Raster Main

Figure 1. Example of Locality Selection Interface.

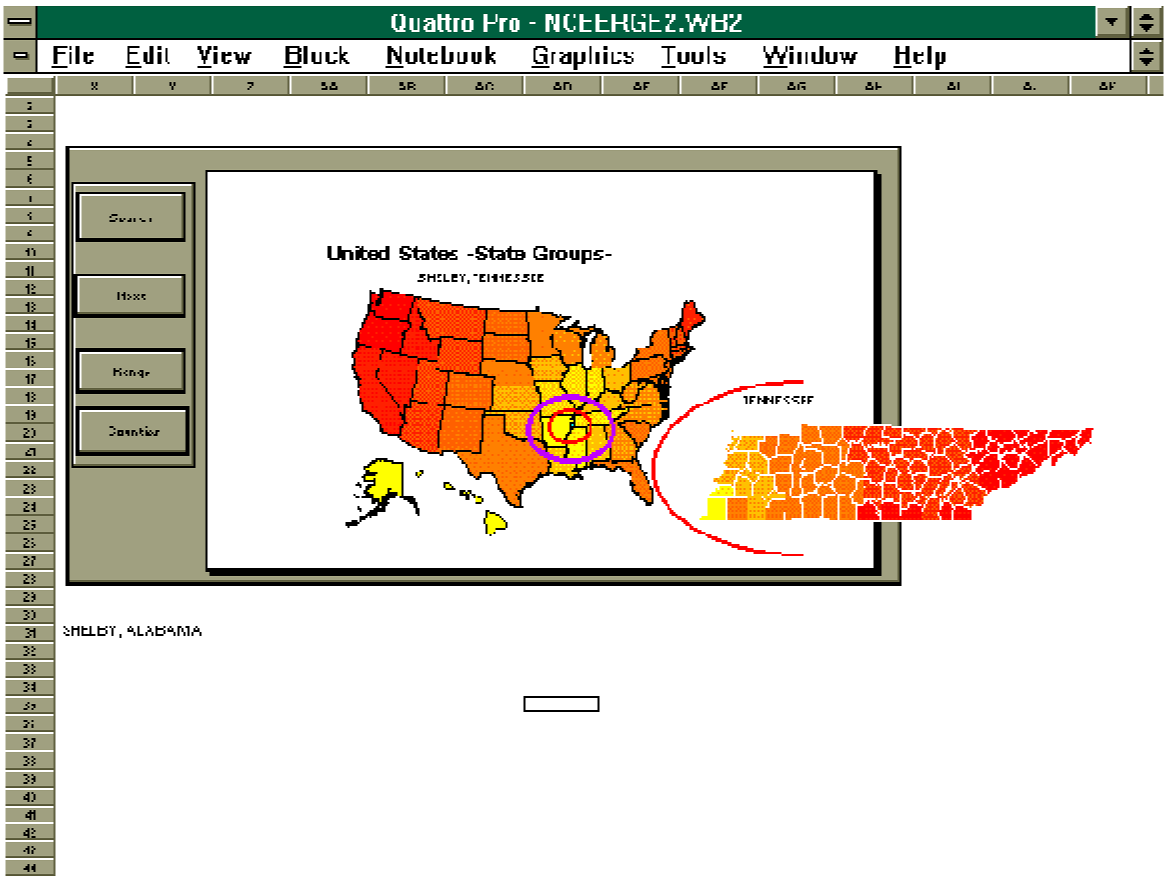


Figure 2. Flow Diagram for Construction and Solution of Model.

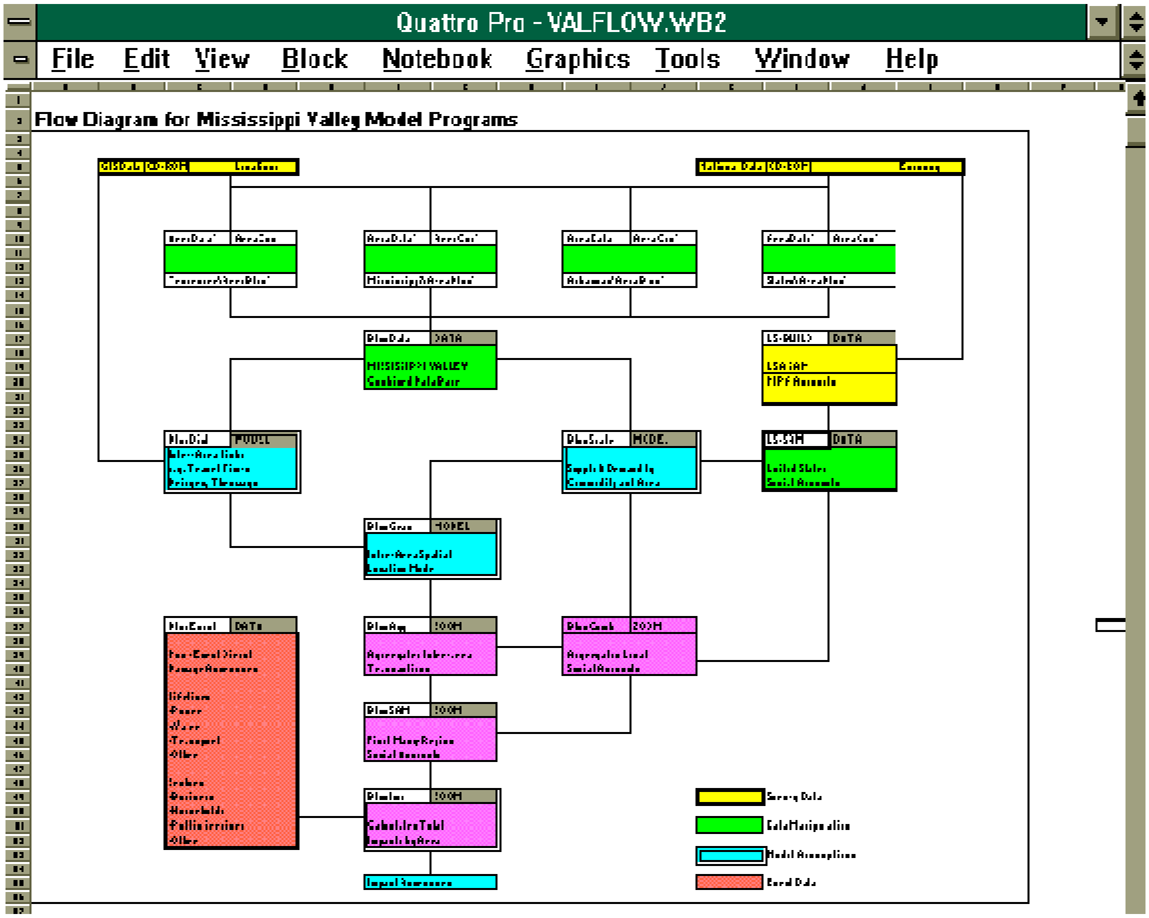

Quattro Pro - VALFLOW.WB2
File Edit View Block Notebook Graphics Tools Window Help
Flow Diagram for Mississippi Valley Model Programs

Figure 3. Example of Many-County Social Accounts

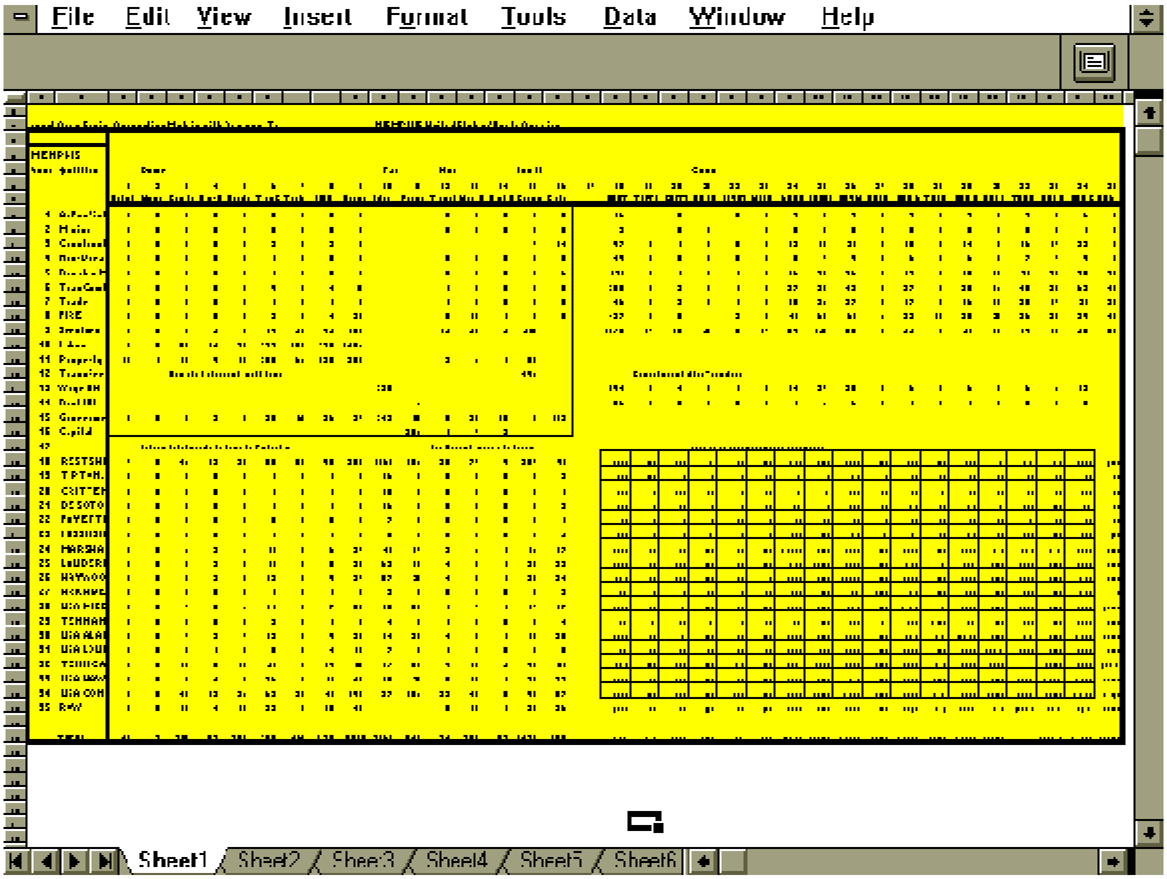
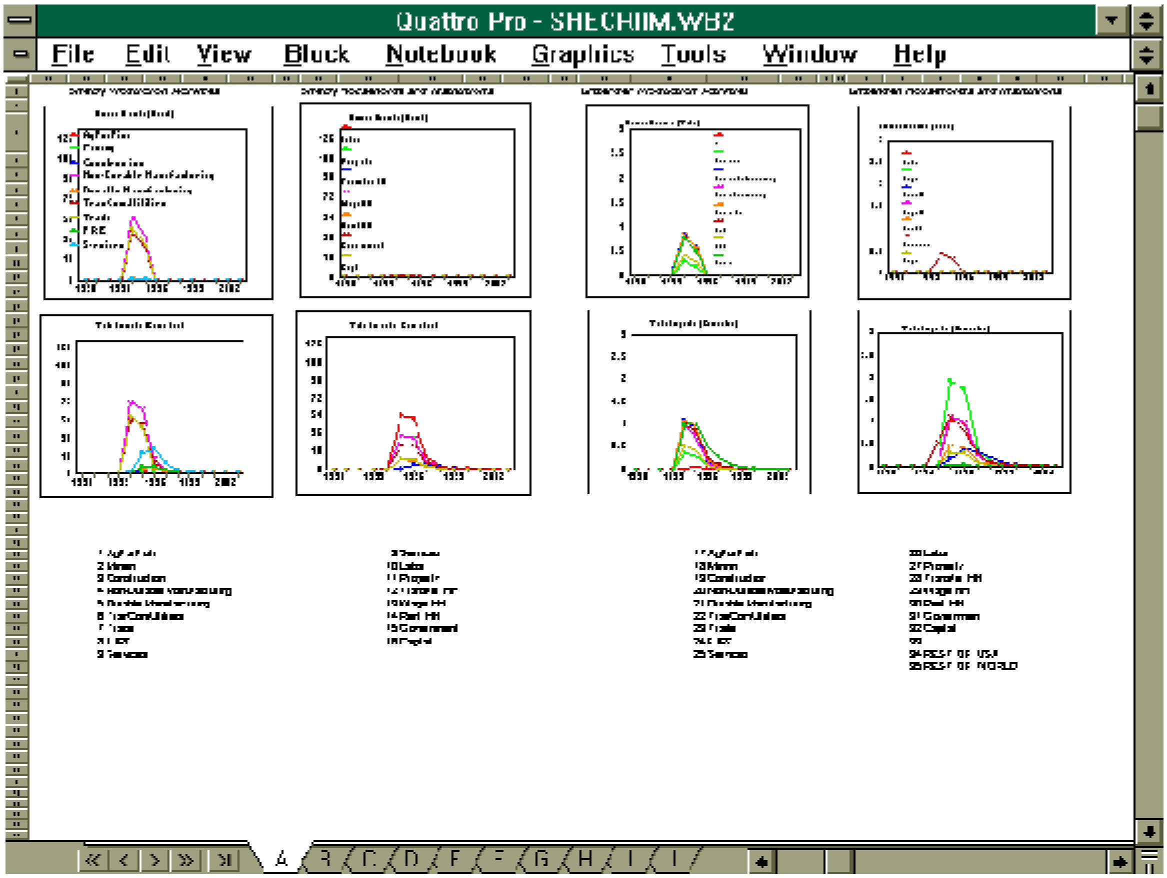


Figure 4. Example of Time-Lagged Impacts


Quattro Pro - SHECHIM.WB2
File Edit View Block Notebook Graphics Tools Window Help

The extraction of a relief index using digital data in the Idrisi GIS

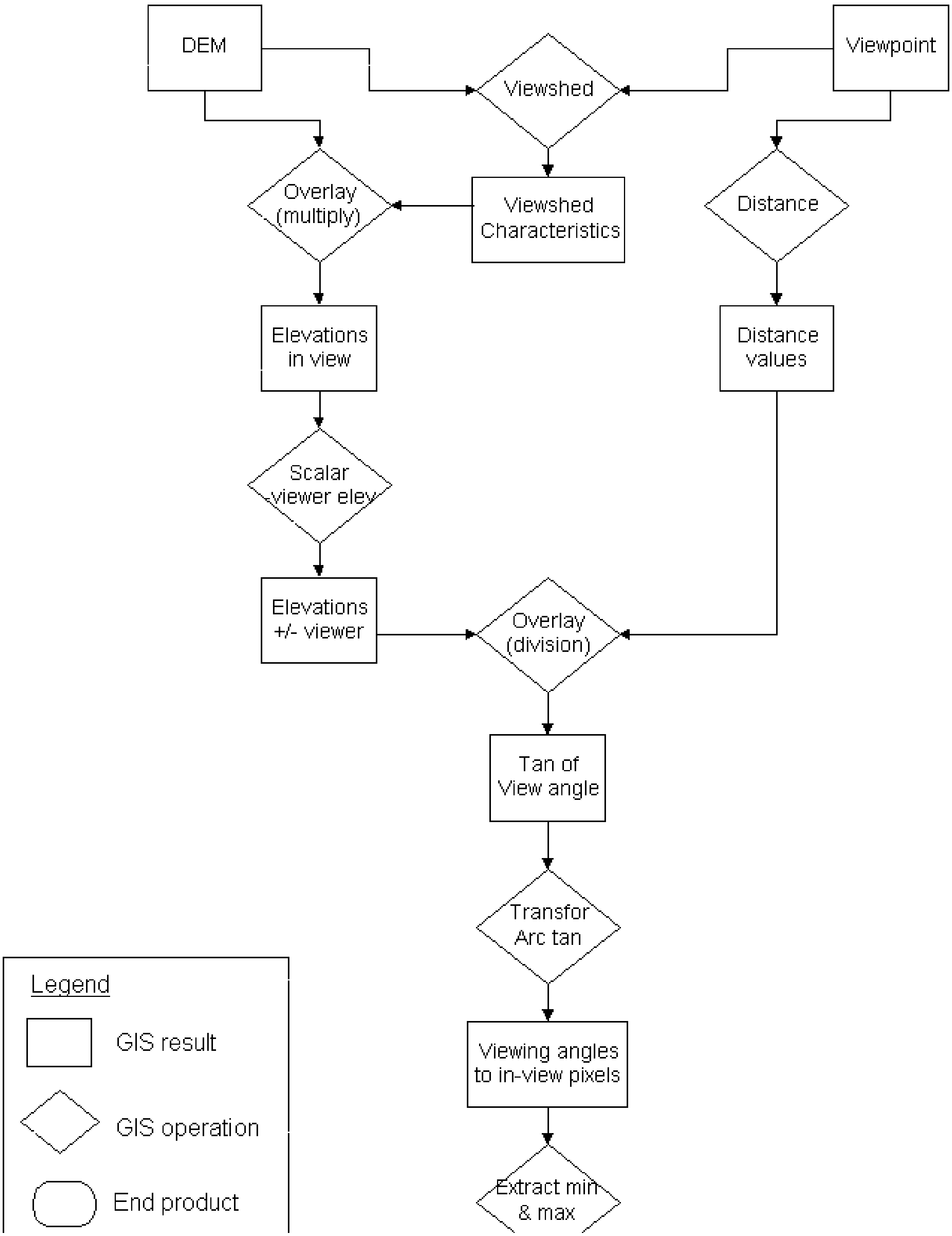


Figure 1.

The extraction of a 'depth of view' index using digital data in the Idrisi GIS

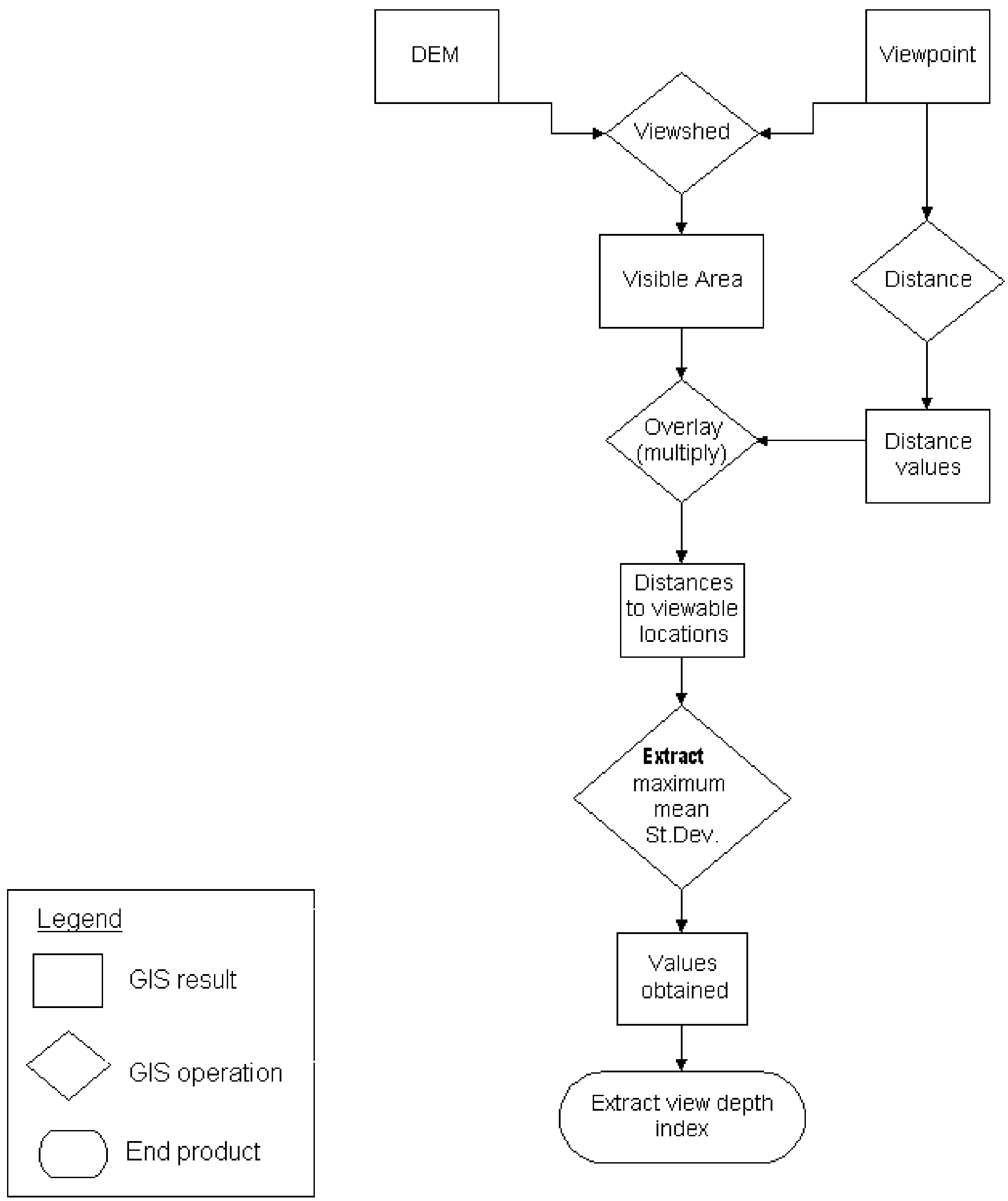
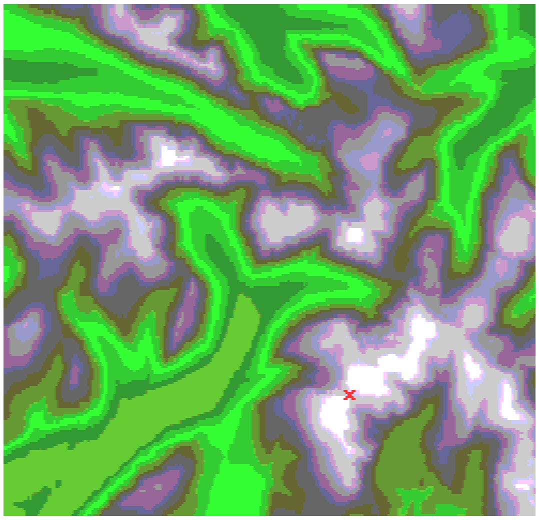
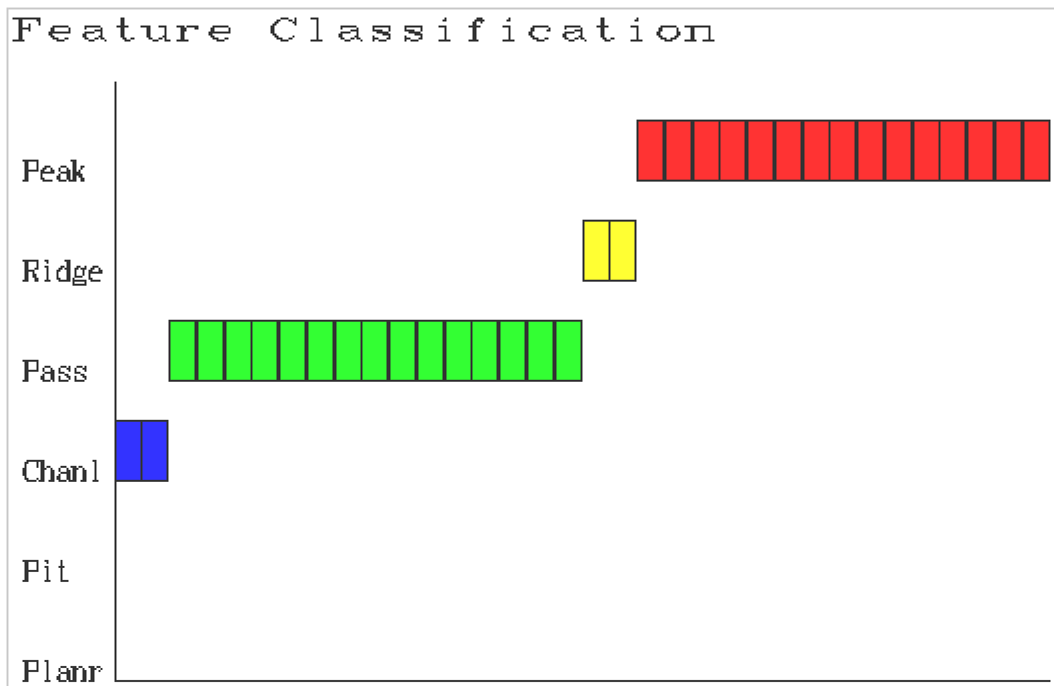


Figure 2.




Feature Classification
Peak
Ridge
Pass
Chanl
Pit
Planr

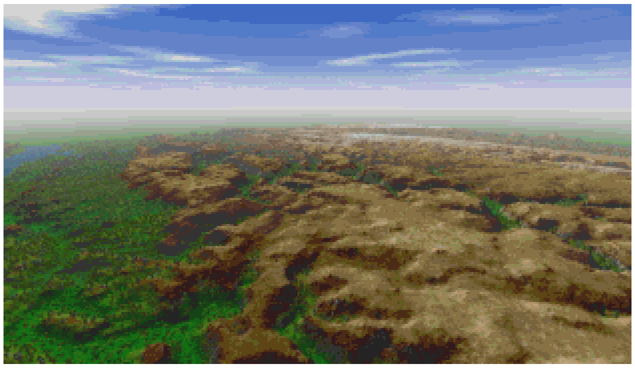


Figure 1. A landscape generated for a portion of the Williamsport East 1:250,000 DEM.

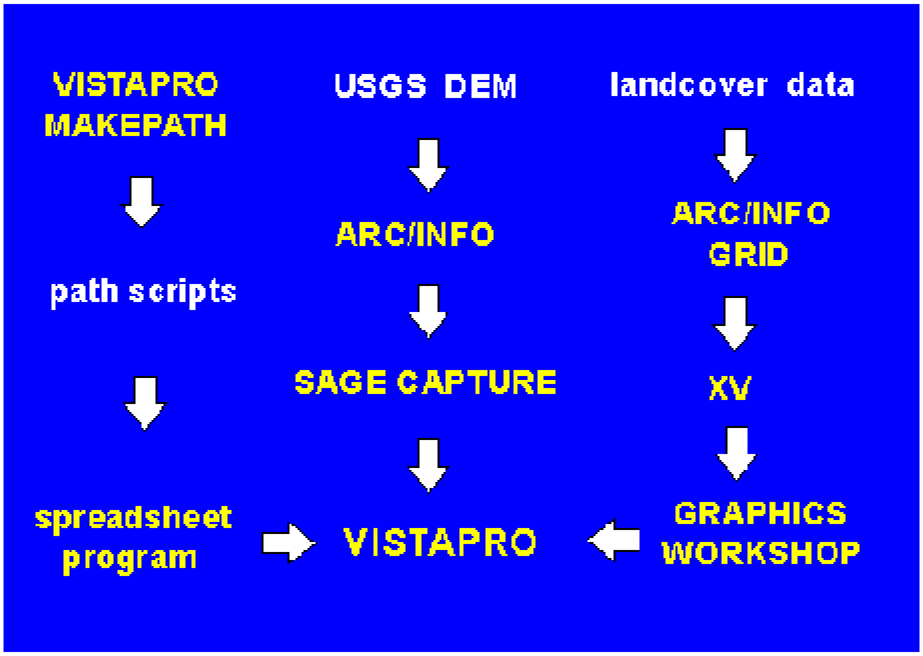
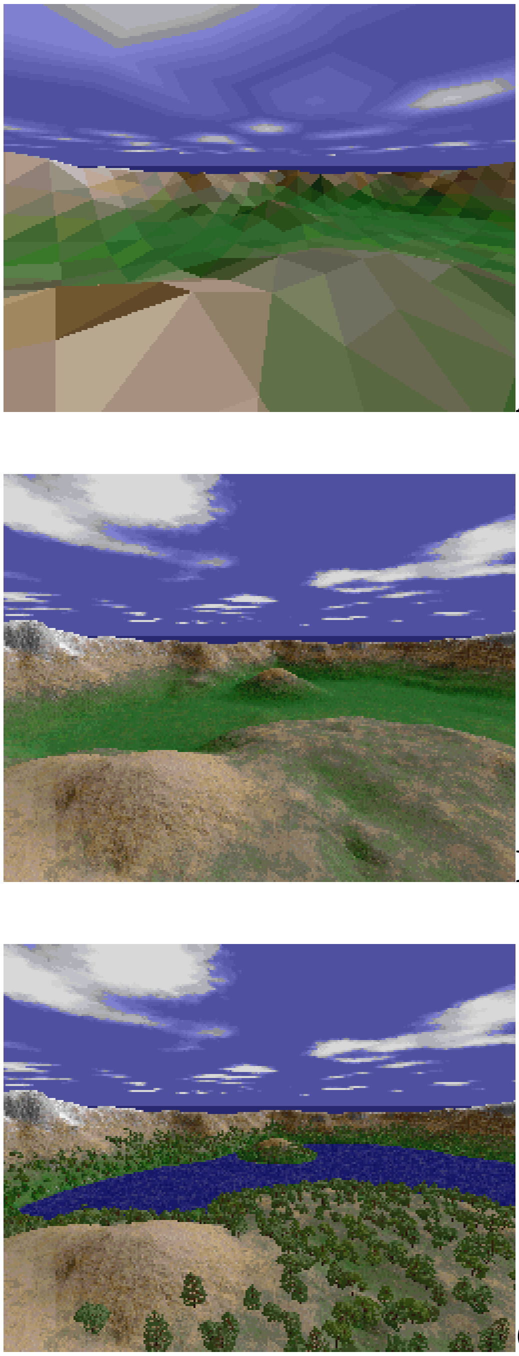


Figure 2. Flow chart of procedures used. White text indicates input data; yellow text indicates software programs.

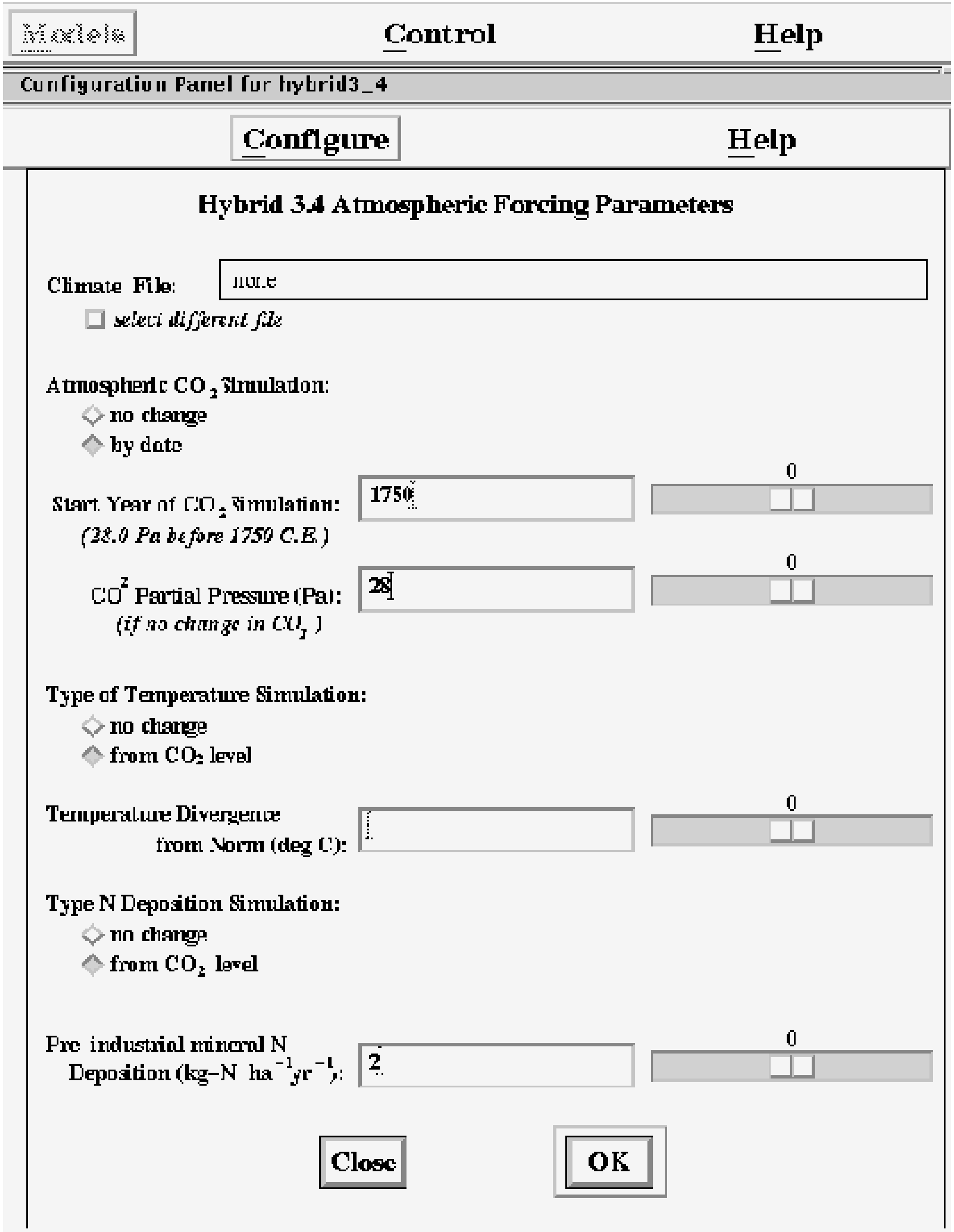


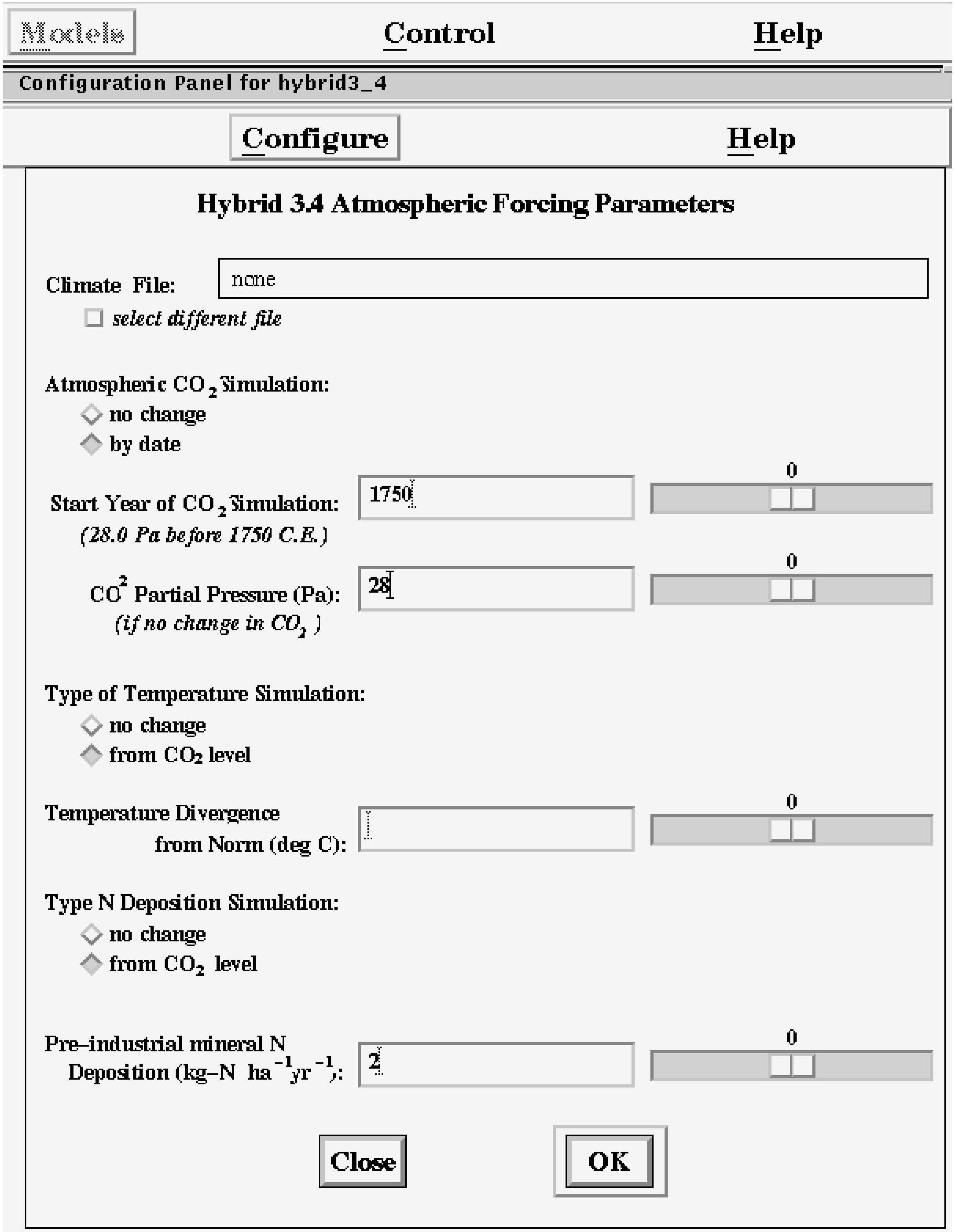
A

B

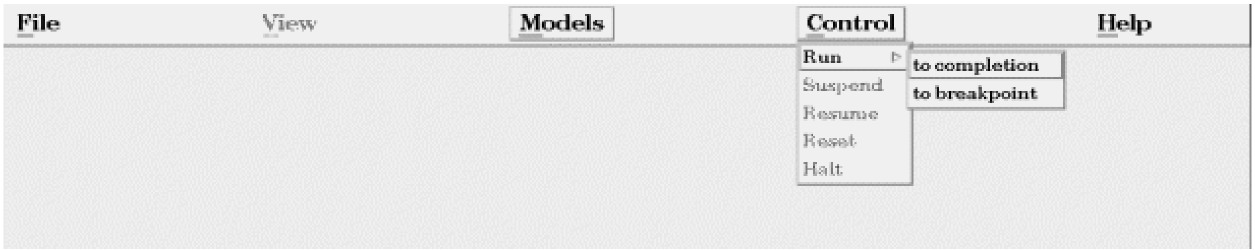
C

Figure 3. Landscape scenes rendered from the 1:24000 Crater Lake DEM. A. Low resolution. B. High resolution. C. High resolution with fractal models of terrain type.


Models
Control
Help
Configuration Panel for hybrid3_4
Configure
Help
Hybrid 3.4 Atmospheric Forcing Parameters
Climate File:
none
select different file
Atmospheric CO2 Simulation:
no change
by date
Start Year of CO2 Simulation:
(28.0 Pa before 1750 C.E.)
1750
0
CO2 Partial Pressure (Pa):
(if no change in CO2)
28
0
Type of Temperature Simulation:
no change
from CO2 level
Temperature Divergence
from Norm (deg C):
0
Type N Deposition Simulation:
no change
from CO2 level
Pre-industrial mineral N
Deposition (kg-N ha-1 yr-1):
2
0
Close
OK


Models
Control
Help
Configuration Panel for hybrid3_4
Configure
Help
Hybrid 3.4 Atmospheric Forcing Parameters
Climate File:
none
select different file
Atmospheric CO2 Simulation:
no change
by date
Start Year of CO2 Simulation:
(28.0 Pa before 1750 C.E.)
1750
0
CO2 Partial Pressure (Pa):
(if no change in CO2)
28
0
Type of Temperature Simulation:
no change
from CO2 level
Temperature Divergence from Norm (deg C):
0
Type N Deposition Simulation:
no change
from CO2 level
Pre-industrial mineral N Deposition (kg-N ha-1yr-1):
2
0
Close
OK


File
View
Models
Control
Help
Run
Suspend
Resume
Reset
Halt
to completion
to breakpoint


File
View
Models
Control
Help
Run
Suspend
Resume
Reset
Halt
to completion
to breakpoint

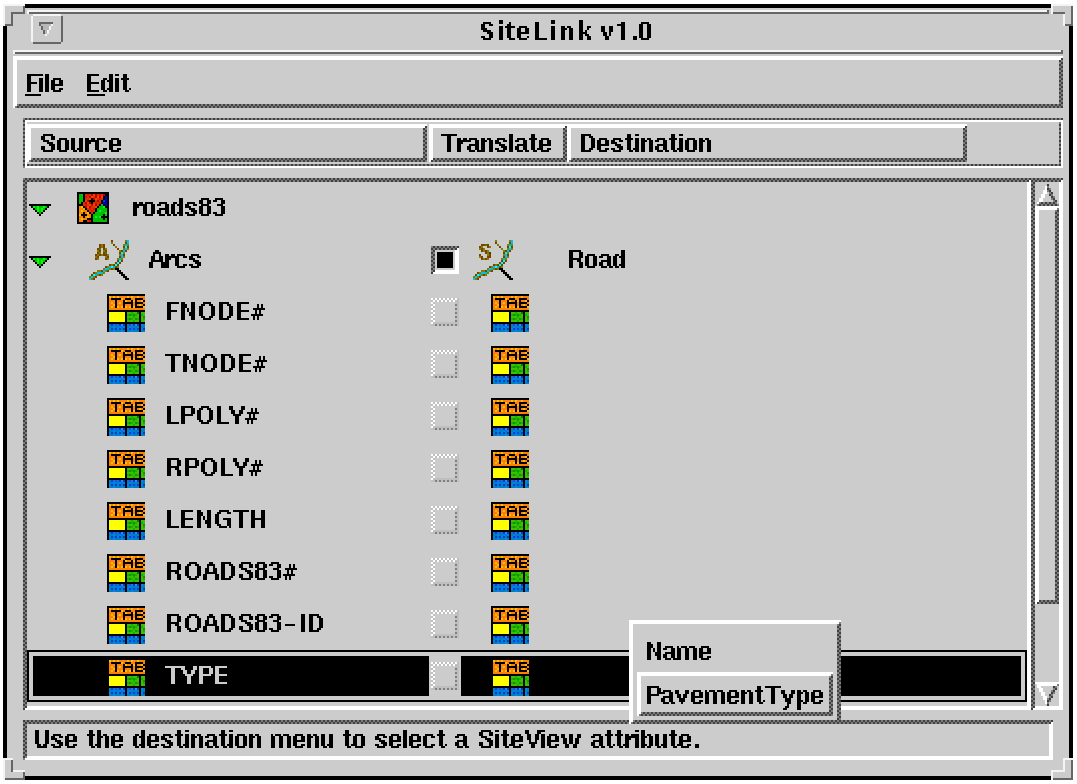


Figure 2. (a) Using the SiteLink translator to translate an ARC/INFO roads data set. The "Road" object type has been assigned, and the PavementType attribute slot is being assigned for the INFO attribute "TYPE."

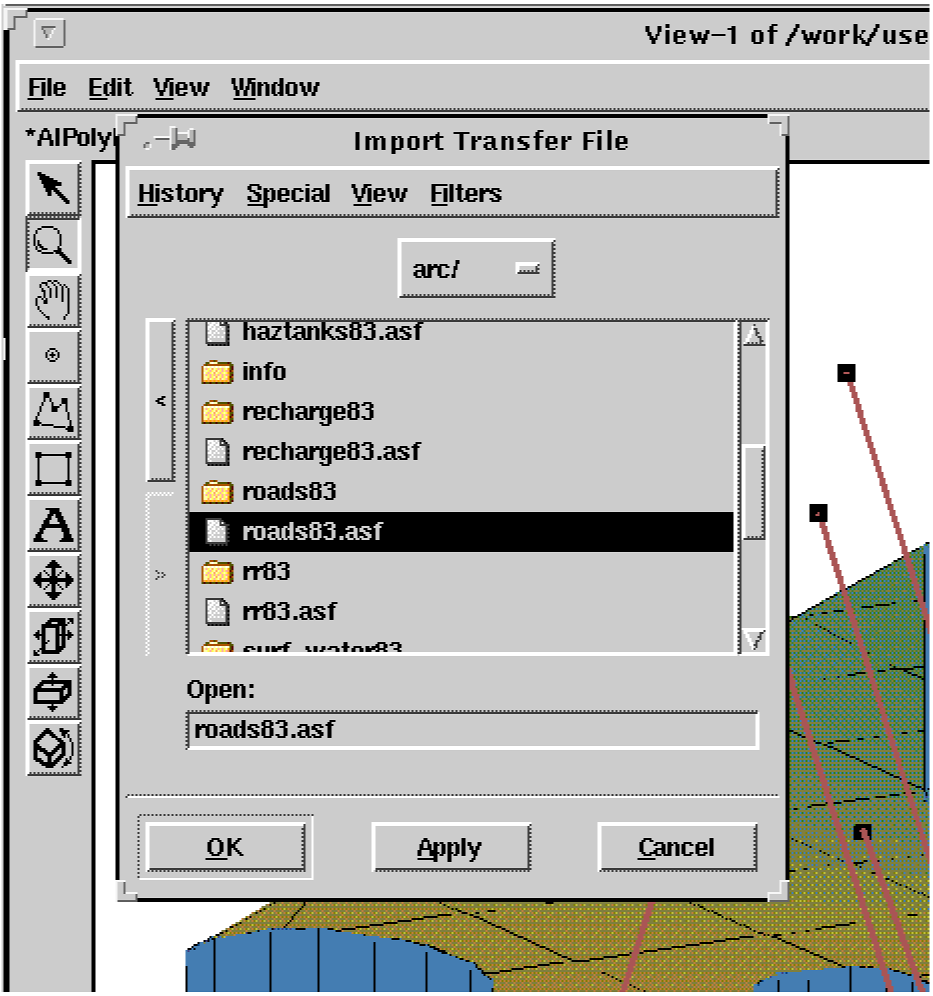


Figure 2. (b) Importing the roads SiteLink transfer file into SiteView.

Figure 3. Base map produced in ARC/INFO with 1962 aerial photograph showing the study area.

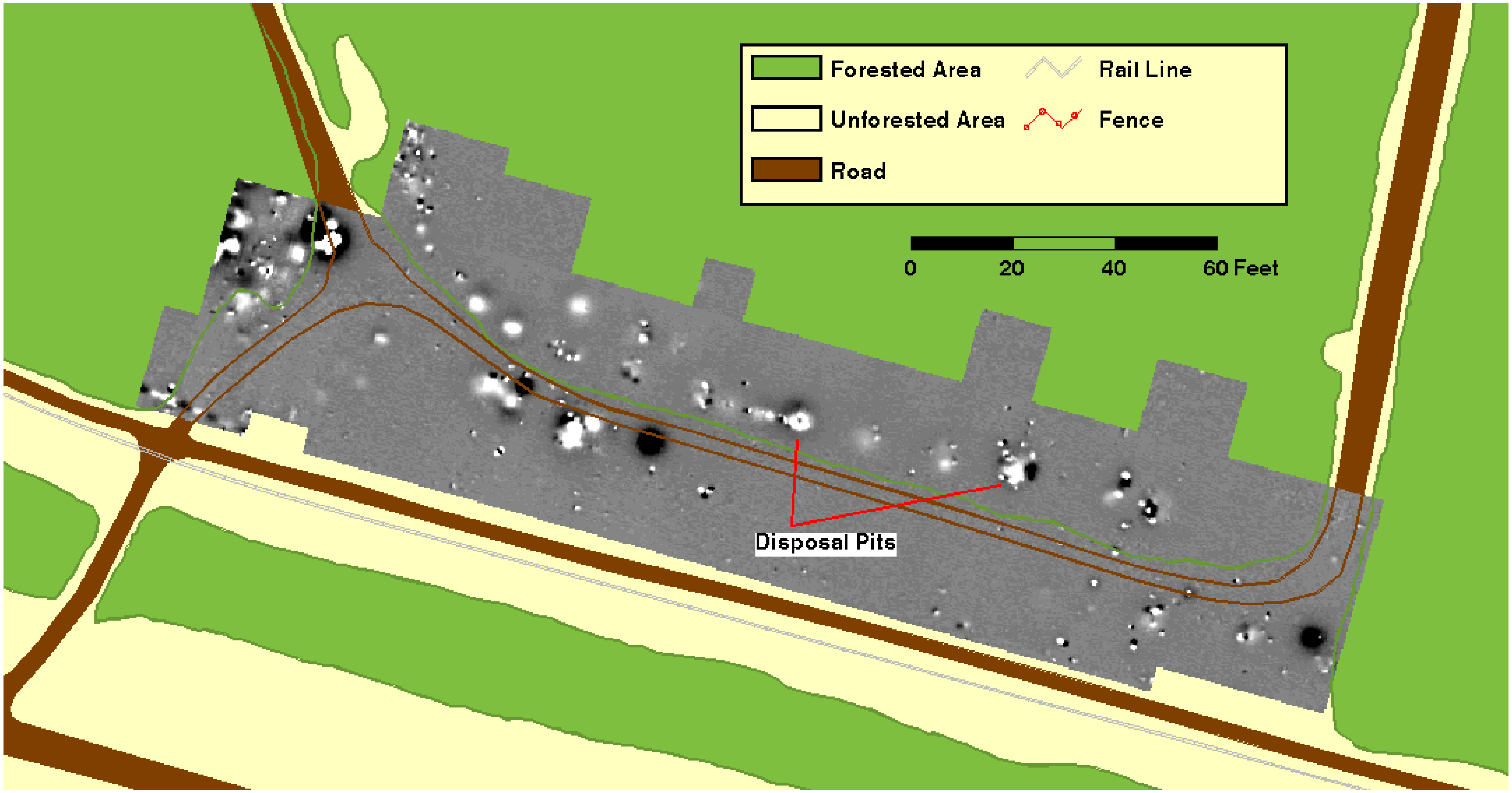


Figure 4. Nonintrusive magnetic gradiometer survey data shown in ARC/INFO.

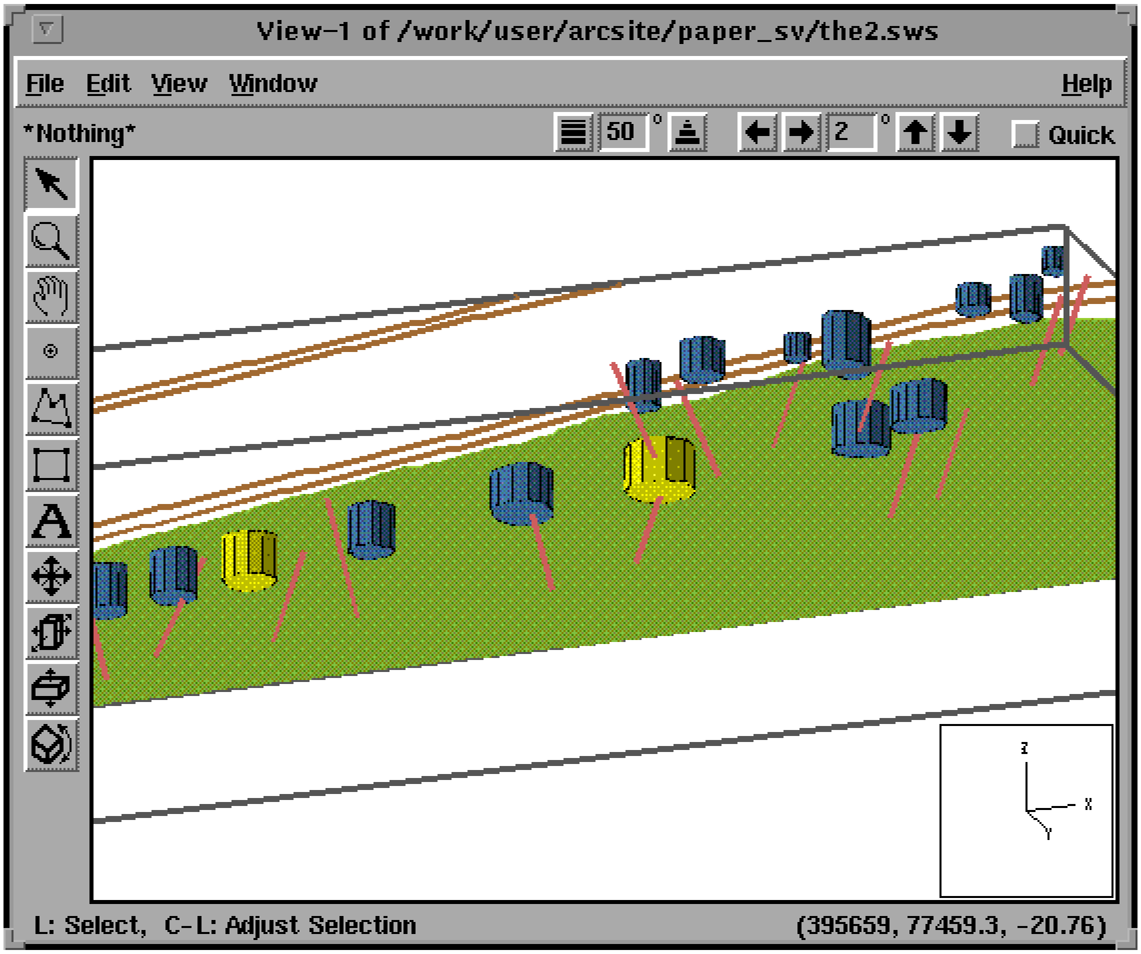


Figure 5. (a) Subsurface view of the study area in SiteView showing disposal pits and slant bores. Surface map layers were translated from ARC/INFO.

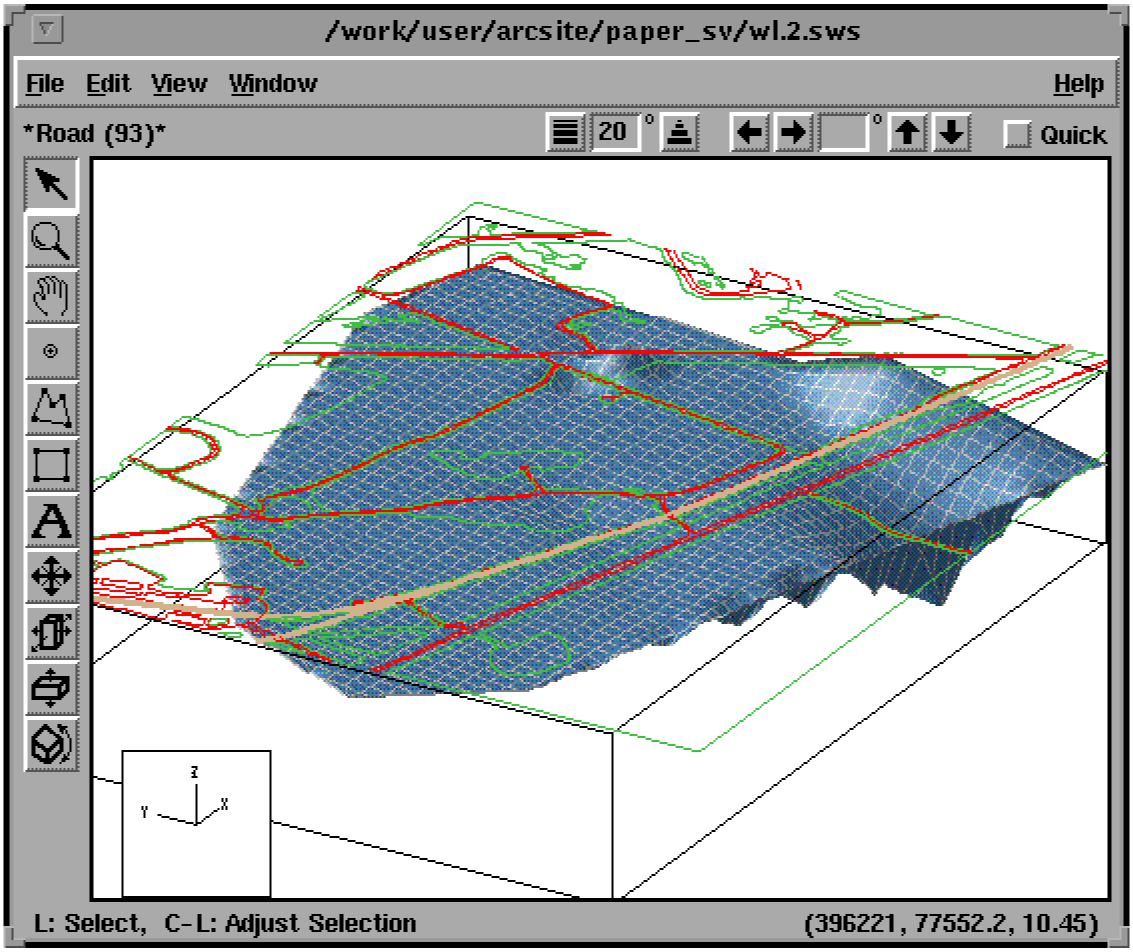


Figure 5. (b) Groundwater depth visualized in SiteView with translated base map layers superimposed on the surface.

1 Introduction

This article presents results from a research project at SINTEF Informatics carried out in 1994 (For a more detailed description of this work, see [MJHH95].) The main objective was to explore the assumed synergy in combining two separate R&D fields: Geographic Information Technology (GIT), and planning and scheduling based on Constraint Reasoning (CR). Forest management was chosen as an interesting and challenging application area. NORSKOG, a Norwegian association of forest owners, provided problem requirements and data needed to design, implement and test a software prototype for decision support in long term harvest scheduling.

During the past decade, the forest trade has faced a set of new challenges, both in Norway and in other parts of the world. Authorities and market segments demand accomplishment and documentation of sustainable forest harvesting. In addition to reaching economical objectives, the trade also has to take care of ecological concerns, such as wildlife preservation and biological diversity. Occasionally, recreational objectives have to be addressed.

Long term treatment schedules are considered one of the main vehicles in documenting adherence to external constraints, and, as control guidance for sustainable forest harvesting. The corresponding scheduling problem, which often is referred to as *spatially explicit* or *stand specific* treatment scheduling, has received considerable attention during the last ten years, both from the forestry research society, software vendors, and the operations research community. Still, there is no software available to support the specific scheduling problems arising in Norwegian forestry. Existing methods do not handle wide planning horizons, or, they are not capable of handling spatial constraints and criteria.

The remainder of the paper is organised as follows: In section 2 we describe and discuss the Clear-Cut Scheduling Problem (CCSP) from a scheduling point of view. We outline some of the traditional approaches and describe the selected strategy. The design and implementation of an experimental prototype called *ECOPLAN* are described in section 3. A specific test case and selected results are presented in section 4. Some final remarks are given in section 5.

Next: 2 The Clear-Cut Scheduling **Up:** Solving the Clear-Cut Scheduling **Previous:** Solving the Clear-Cut Scheduling

---

gunnar.misund@si.sintef.no

2 The Clear-Cut Scheduling Problem (CCSP)

We define the *Clear-Cut Scheduling Problem (CCSP)* as the assignment of clear-cut years to individual treatment compartments in a forest area. Given a forest area and its subdivision in stands, the task in CCSP is to generate a harvesting schedule for a given horizon which a) satisfies a number of constraints, and, b) strikes a careful balance between several criteria. In the following sections, we shall discuss the successful solution of an instance of the CCSP based on Norwegian legislation in general, and, in particular, restrictions on forest harvesting in the near-city areas of Oslo.

2.1 Description

The goal of the CCSP is to find a *complete*, *consistent*, and *optimised* clear-cut harvesting schedule. By *complete* we understand that each region must be assigned a time for future treatment. By *consistent* we mean that the schedule must not violate any hard constraints (A hard constraint is a relation which must be satisfied. A soft constraint may be relaxed). By *optimised*, we understand that the solution must optimise on certain criteria, e.g., by minimising or maximising the value of a defined objective function, and, it must satisfy all soft constraints to as large degree as possible. Time granularity is one year.

Topological description. The considered forest area is completely partioned into a number of non-overlapping regions. The underlying assumption is that each region is homogeneous with respect to the forest properties that are relevant to harvest scheduling (i.e., regions are stands).

Individual parameters/functions. For every region, several parameters/functions are given: The time of the most recent harvesting, minimum duration between harvests, maximum duration between harvests, optimal time between harvests (age/ripeness), the time it takes for trees to grow from 0 to a certain height, the area of each region, and the volume that may be harvested a number of years after the last harvesting.

Hard constraints. Before a region may be harvested, it is required that every neighbouring region has an average tree height of say at least 2 meters. We shall denote these hard constraints the *2-m constraints*.

Soft constraints. For economic and quality reasons, there are bounds on times between harvesting. These constraints may be relaxed in order to fulfill the 2-m constraint.

Criteria for an optimised schedule. Below we describe the four major optimisation criteria identified by forestry experts for the CCSP.

Optimal Harvesting Time. For every region, the harvesting time should be as close as possible to its optimal harvesting time.
Even Consumption. The estimated harvesting volumes for each year should be as close as possible to the average harvested volume.
Old Forest. The schedule should maintain a minimum area of forest above a given age threshold, over the schedule horizon. We may, for instance, want to minimise the sum of violations of the old forest constraint.
Visual Impact. The schedule should minimise visual damage relative to a given set of viewpoints. By projecting the landscape (requiring a terrain model) to the respective viewing frames, it is possible to calculate the total area in these images which correspond to clear-cuts. We then may want to minimise the maximum clear-cut contribution (the ``worst" visual impact) over all years.

2.2 Problem Solving Techniques

The CCSP is an example on a complex combinatorial optimisation problem. Focusing on the hard 2-m constraint, the CCSP may be regarded as a *Constraint Satisfaction Problem (CSP)* [Tsa93]. In particular, there are strong similarities between the CCSP and the *Graph Colouring Problem (GCP)*. Informally, the task in the GCP is to assign colours (from a given set of colours) to the nodes in a graph in such a way that no neighbours is given the same colour. The GCP belongs to the class of **NP**-complete problems [GJ79], for which there probably does not exist any efficient (polynomial) algorithm. Although there are additional constraints and objectives, our conjecture is that the CCSP is **NP**-hard (A proof is beyond the scope of this paper). We must therefore lower our expectations and concentrate on finding high-quality solutions in limited time. Adding the complexity, this problem indeed calls for efficient, robust and flexible optimisation techniques.

Mathematical Programming. *Linear Programming*, in particular, *Mixed Integer Programming (MIP)* and *Goal Programming (GP)* have been applied to the CCSP and similar problems [WMMK94]. Our initial attempts to formulate the CCSP as a MIP has lead us to conclude that this approach is not well suited, for the following reasons: 1) lack of flexibility in expressing constraints and objective criteria, 2) lack of support for mixed-initiative problem solving and 3) lack of repositories for heuristics to guide combinatorial search.

Systematic Tree Search (STS). Taking a Constraint Satisfaction Problem perspective on the CCSP, several backtracking tree search and consistency techniques are viable. *Standard Backtracking (SB)* may be seen as the basis for these techniques. SB will iteratively construct a solution by successively assigning values to the problem variables (i.e., assign a harvesting year to one region, then to another, and so on) while checking whether constraints are satisfied. If no value is possible for the current variable due to constraint violations, the algorithms will backtrack and try a new value for the previously instantiated variable. We have evaluated STS with several variable and value ordering heuristics in the context of finding a 2-m feasible solution. In initial empirical investigations it was not possible to obtain a solution within acceptable response limits, even for small CCSP problem instances (due to the exponential time complexity of STS).

Iterative Improvement Techniques (IIT). Over the past few years, these methods have shown remarkable performance in providing high quality solutions to scheduling problems in limited time [Dor95]. The basic idea is *neighbourhood search*, i.e., given any complete solution, generate a neighbourhood by applying a set of modification operators. The search for a better solution then proceeds iteratively by selecting the best neighbour as the new current solution. In this basic form, IIT is a hill-climbing algorithm which might get stuck in local optima. To remedy this, so-called *meta-heuristics* may be employed, e.g., *Simulated Annealing (SA)* [Kir83] or *Tabu Search (TS)* [Glo90]. A particularly nice feature of IIT in the context of decision-support is their *anytime* characteristic. The iterative problem solving process may be interrupted at any time, and the best solution so far is available for presentation. IIT has earlier been applied to forest management problems, e.g., to solve the afforestation problem[MJTVV92].

2.3 IIT - The Selected Search Strategy

Our selected search strategy for the CCSP is IIT with the Tabu-Search (TS) meta-heuristic [Glo90]. TS is composed of a neighbourhood operator, an evaluation function for neighbours, a tabu criterion, and an aspiration criterion. A method for generating an initial candidate solution is needed to initiate the iterative improvement process.

The Evaluation Function assigns goodness values to candidate schedules. We have selected a straightforward approach where the evaluation function is a weighted sum of the four optimisation criteria components described above. In addition, a penalty function for violations of the 2-m constraints is introduced as a component. The selection of appropriate weights is non-trivial.

Neighbourhood Operator. We have selected a neighbourhood operator which simply generates the neighbourhood by modifying exactly one harvesting year. The operator generates only local-feasible harvesting years (i.e. within the legal harvesting year interval) relative to the selected modified region. This operator is simple, but generates a large neighbourhood.

Initial Schedule. We have selected a greedy algorithm for generating the initial schedule. It assigns the local-optimal harvesting time to every region, where possible.

The Tabu Criterion specifies moves that are tabu and thus will not be executed. In TS, the iterative improvement basically consists of movement to the neighbour with the best value of the evaluation function. To escape from local optima, neighbours with certain defined properties are defined as tabu. Currently, we use a simple criterion, stating that we are not allowed to move to a neighbour which harvesting year has changed within a certain number of iterations.

The Aspiration Criterion. In TS, a move which is defined as tabu may be performed if allowed by an aspiration criterion. Our current choice of aspiration criterion checks for global improvement. If a move is deemed tabu, but will result in the best schedule encountered so far, the move will be performed anyway.

gunnar.misund@si.sintef.no

3 Design and Implementation of ECOPLAN

The ECOPLAN prototype is designed as a synthesis of four modules. The core of the system is a close integration of a scheduling engine called the *IIT Kernel* and a set of *GIT Services*. These two modules are embedded in an interface environment to facilitate interaction with the operator and communication with data sources. The functionality of the four components is briefly outlined below.

User Interface: Information on, and access to, input data. Control of the optimisation process (parameter settings, manual interruptions). Modification of existing constraints/criteria and addition of new ones. Selections of output presentations, including animations. Currently implemented exclusively with development and evaluation of the IIT Kernel and the GIT Services in mind, with a parameter file as the main control.

Data Interface: Communication with external data sources. Conversion of input data. Output of results. Currently taken care of by means of plain ASCII files.

GIT Services: Generation of customised terrain models from scattered data, such as elevation contours, 3D data on road and stream networks, and geodetic points. Generation of consistent topological models from spaghetti data of site polygons. Spatial calculations of for example area, perimeter and distance. Visual viewpoint analysis. Preparation of colour coded digital maps. Preparation of data for 3D visualisation.

The GIT services are composed of a diversity of public domain tools and a suite of spatial analysis methods designed and developed particularly for the ECOPLAN module. The terrain generation and analysis, e.g., the Visual Impact analysis, is implemented by extensive use of the SISCAT library [sis95], [ADH95], a comprehensive, C++ based toolkit for construction of surfaces from various kinds of scattered data.

IIT Kernel: Modeling and management of all information relevant to the scheduling optimisation. Methods and algorithms for iterative schedule improvement. Implemented from scratch in the object-oriented language C++, in a way which makes it easy to plug in new iterative improvements techniques (see section 2.2), and to implement new constraints and criteria.

Next: 4 A Case and **Up:** Solving the Clear-Cut Scheduling **Previous:** 2 The Clear-Cut Scheduling

---

gunnar.misund@si.sintef.no

4 A Case and Some Results

To enable the investigation of behaviour and performance of the ECOPLAN prototype, we were provided with data on a forest area in the South-Eastern part of Norway.

Our test case is a 16 sq.km. forest, see figure 1, of which roughly speaking 85% is considered productive. The area is subdivided into approximately 500 stands. The average size of a stand is 28.000 sq.m., varying from 200 sq.m. to 148.000 sq.m. The forest is relatively young, about 60% of the total area consists of forest which is less than 30 years old.

[IMAGE]

The ECOPLAN prototype handles site-specific information. Some simplifications have been made:

- All stands consist of only one single wood species.
- All stands are considered equal with respect to site quality.
- The volume growth function is designed as simply as possible, i.e., linear growth up to a given threshold age, and stagnation thereafter.
- The only forest treatment considered is clear-cutting.
- ``Old'' forest is defined to be more than 60 years for all stands.

The scheduling horizon is 100 years. For each year, we need to determine the set of stands to be treated. In the next sections we present results based on a series of experimental runs of ECOPLAN with different parameter settings. The presentation focuses on one single constraint or criterion at a time. However, the results are generated with all constraints and criteria simultaneously active.

4.1 2-m constraint.

The 2-m constraint implies that all neighbouring stands to a clear-cut region must be higher than 2 meters. Experiments showed that all 2-m constraints are satisfied after a relatively low number of iterations. Typically, the number of 2-m constraint violations decreases dramatically during the first 100 iterations. It takes equally many iterations to resolve the last few conflicts in a typical case.

Even Consumption. The Even Consumption criterion implies that the optimisation process seeks to distribute harvesting volume evenly over the scheduling period. In figure 2, we illustrate how this criterion gradually improves. Harvested volume as a function of time over the scheduling horizon is shown for three different schedules, the initial greedy solution, and the 600 and 1200 iteration schedules, respectively.

[IMAGE]

The initial solution is, as described in section 2.3 and 4.1, generated by cutting the forest at the locally optimal harvest time. Due to the structure of the input data (the stands are classified according to five year intervals), we get large consumed volumes every 5th year and no activity in the intermediate years. The schedule is considerably improved after 600 iterations, and after 1200 rounds the yearly consumption seems to stabilise at an approximately even level.

Old Forest. Due to ecological considerations, the area percentage of old forest should be kept above a certain threshold level. In our case, we define old forest to be 60 years or older, and we want 40% or more of the total area to be occupied by stands of this age class.

The initial stage of the case forest is characterised by a large amount of young stands. Thus, the forest needs time to meet the old forest criterion. In the greedy solution, stands are cut as soon as they reach optimal ripeness which is above the old forest age. Hence, we get an uneven, oscillating pattern.

This pattern is considerably improved during the optimisation process. In an optimised schedule, we reach a stable situation after approximately 45 years of harvesting. The desired level of old forest is reached, and the oscillating behaviour is avoided.

Visual Impact. In figure 3, we have simulated the visual impact from a given viewpoint in a certain year of a good (right) and bad (left) schedule. In the 3D views, clearings and young stands are rendered light grey, while older forest is darker. The left view presents a landscape which may be characterised as totally demolished, while the right visual impact is satisfactory.

[IMAGE]

The optimisation procedure will favour a schedule were many of the vistas are visually pleasant. Solutions that contain visually unpleasant harvesting patterns are heavily punished.

Optimal Harvest Time. The initial solution is defined as a schedule where the Optimal Harvest Time criterion is maximised, i.e. that every stand is clear-cut exactly in the year when the forest is considered to be as ``ripe'' as possible. In contrast to the other criteria, the Harvest Time goodness will inevitably decrease during the optimisation process when all constraints/criteria are active. After 1200 iterations, we reach an acceptable result. Approximately 25% of the regions are still optimally harvested. Many stands are clear-cut at near-optimal time. Relatively few stands are found in the sub-optimal parts of their feasible interval.

4.2 Performance

The ECOPLAN prototype is implemented in a UNIX environment, and the tests have been conducted on standard mid-range workstations.

To establish the values of the various parameters, an optimisation session typically started with test-runs with a low number of iterations, and perhaps with only one or two active constraints. Finally, an optimisation process involving a high number of iterations was executed to gain a satisfactory solution.

One single pass involves a preprocessing phase (see section 3) and a number of improvement steps. The preprocessing typically lasts from 20 to 30 seconds up to a few of minutes, depending on the input data. The ECOPLAN prototype generates from ten to hundred new schedules per minute. Not surprisingly, the performance has shown to be relative to the number of active constraints. However, there is a significant potential in optimising the implementation.

Next: 5 Final Remarks **Up:** Solving the Clear-Cut Scheduling **Previous:** 3 Design and Implementation

gunnar.misund@si.sintef.no

5 Final Remarks

The ECOPLAN Project. As presented in section 4, the scheduling strategy for the CCSP reported in this paper has shown excellent performance on real data. Based on the ECOPLAN prototype, an R&D project was launched in September 95. The main goal is to develop a module for long term treatment scheduling under economical and ecological constraints and criteria, customized for Norwegian forestry.

The project will undertake comparative experiments with alternative IIT meta-heuristics as well as other search techniques. The goal is to develop search strategies that are able to handle CCSPs with a number of regions which is an order of magnitude larger than in the case data described in this paper.

Major efforts will be devoted to the refinement of the underlying model from a forestry point of view. Integration with stand simulation software, and the accommodation of several types of treatment (thinning, sparse cutting) are important issues in this context. Special attention will be paid to develop a simple, intuitive, and efficient user interface. Flexible and efficient methods for integration with external information repositories will be provided by implementing an advanced software integration platform.

GIT and Constraint Reasoning. The extra efforts needed to establish and maintain the multi-disciplinary profile of the project proved to be highly rewarding, and a necessary condition to achieve our goals. The supply of well-known methods and techniques for management and analysis of spatial information is large and varying. This is also true for planning and scheduling. However, by combining the technologies, we achieved results beyond expectations at a relatively modest cost.

We suggest that generalised versions of the CCSP problem, such as area planning in the municipalities, agricultural management, campaign planning in marketing and advertising etc., may be solved by similar approaches.

Next: 6 Acknowledgements **Up:** Solving the Clear-Cut Scheduling **Previous:** 4 A Case and

gunnar.misund@si.sintef.no

6 Acknowledgements

We would like to thank NORSKOG for their supply of the Clear-Cut Scheduling Problem and access to real-life case data. We also are in debt to our fellow researchers Trond Vidar Stensby and Per Kristian Nilsen for their competent contributions in the implementation process of the ECOPLAN prototype. Jørgen Haukland at Gjøvik College provided valuable input regarding the the forestry issues of the problem. Finally we want to thank Torgrim Johan Castberg for permitting the use of the forest case data in the SINTEF research project, and Arne Løkketangen at Molde College for introducing us to Tabu Search and providing us with valuable comments and suggestions for improvement.

This work has been partially supported by the Norwegian Research Council, project MOI.31386, and SINTEF Informatics, project 3391 1000.

Next: References **Up:** Solving the Clear-Cut Scheduling **Previous:** 5 Final Remarks

---

gunnar.misund@si.sintef.no

References

ADH95

Erlend Arge, Morten Dæhlen, and Øyvind Hjelle. Mathematical Software for Terrain Modelling. In Jan Terje Bjørke, editor, *SCANGIS '95 - Proceedings of the 5th Research Conference on GIS, 12th 14th June 1995, Trondheim, Norway*, May 1995. ISBN 82-993522-07.

Dor95

Jürgen Dorn. Iterative Improvement Methods for Knowledge-Based Scheduling. *AI Communications*, 8(1):20 -- 34, 1995.

GJ79

M R Garey and D S Johnson. *Computers and Intractability*. Freeman, 79.

Glo90

F Glover. Artificial Intelligence, Heuristic Frameworks and Tabu Search. *Managerial and Decision Economics*, 11:365--375, 90.

Kir83

S Kirkpatrick. Optimization by Simulated Annealing. *Science 220*, 83.

MJHH95

G. Misund, B. S. Johansen, G. Hasle, and J. Haukland. Integration of Geographical Information Technology and Constraint Reasoning - A Promising Approach to Forest Management. In Jan Terje Bjørke, editor, *SCANGIS '95 - Proceedings of the 5th Research Conference on GIS, 12th 14th June 1995, Trondheim, Norway*, May 1995. ISBN 82-993522-07.

MJTVV92

R Mørk Jørgensen, H R Thomsen, and R V Valqui Vidal. The Afforestation Problem: A Heuristic Method Based on Simulated Annealing. *European Journal of Operational Research 56*, pages 184--191, 92.

sis95

SISCAT - The SINTEF Scattered Data Library (version 2.1). Technical report, SINTEF Informatics, Oslo, 1995. Reference manual.

Tsa93

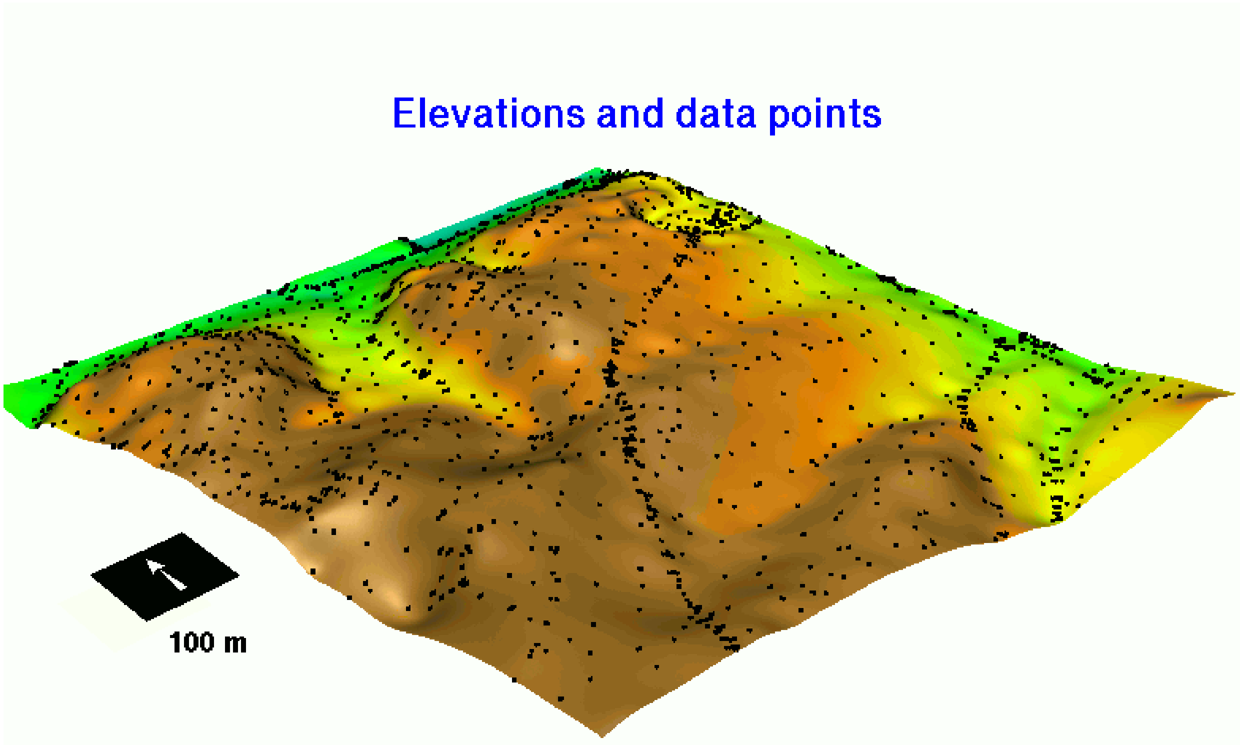
E Tsang. *Foundations of Constraint Satisfaction*. Harcourt Brace & Co, 93.

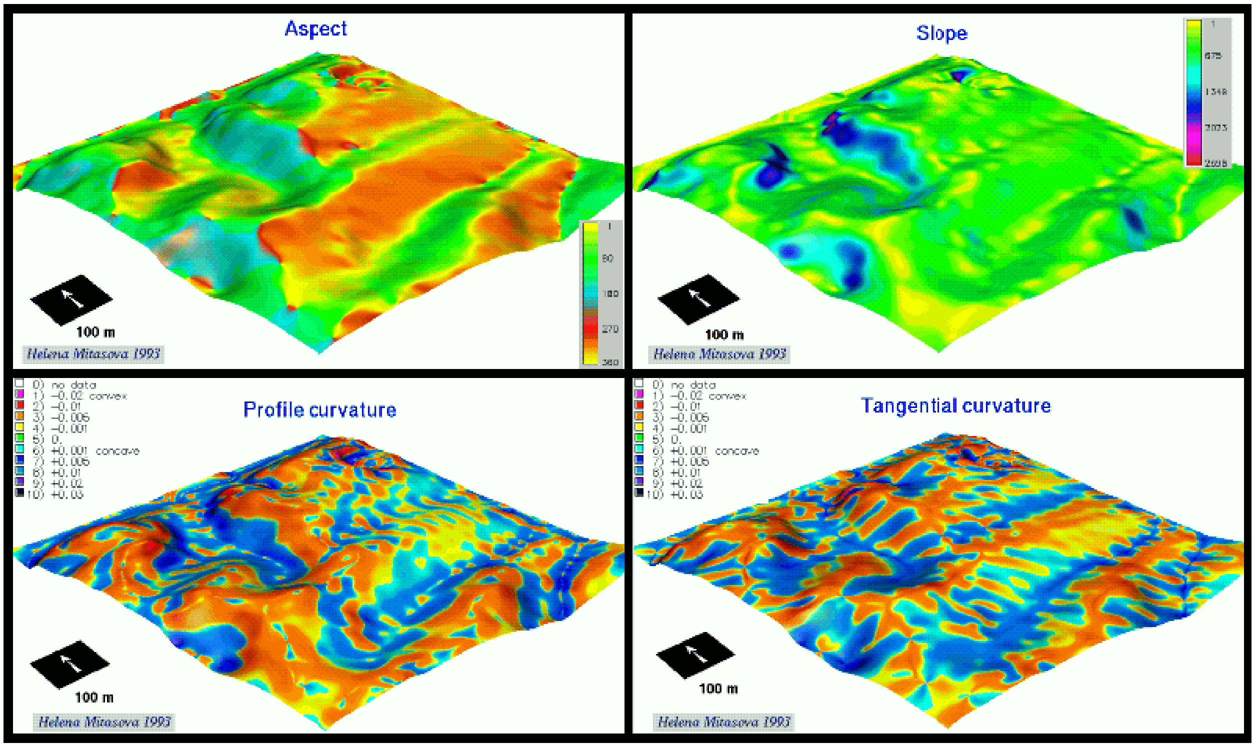
WMMK94

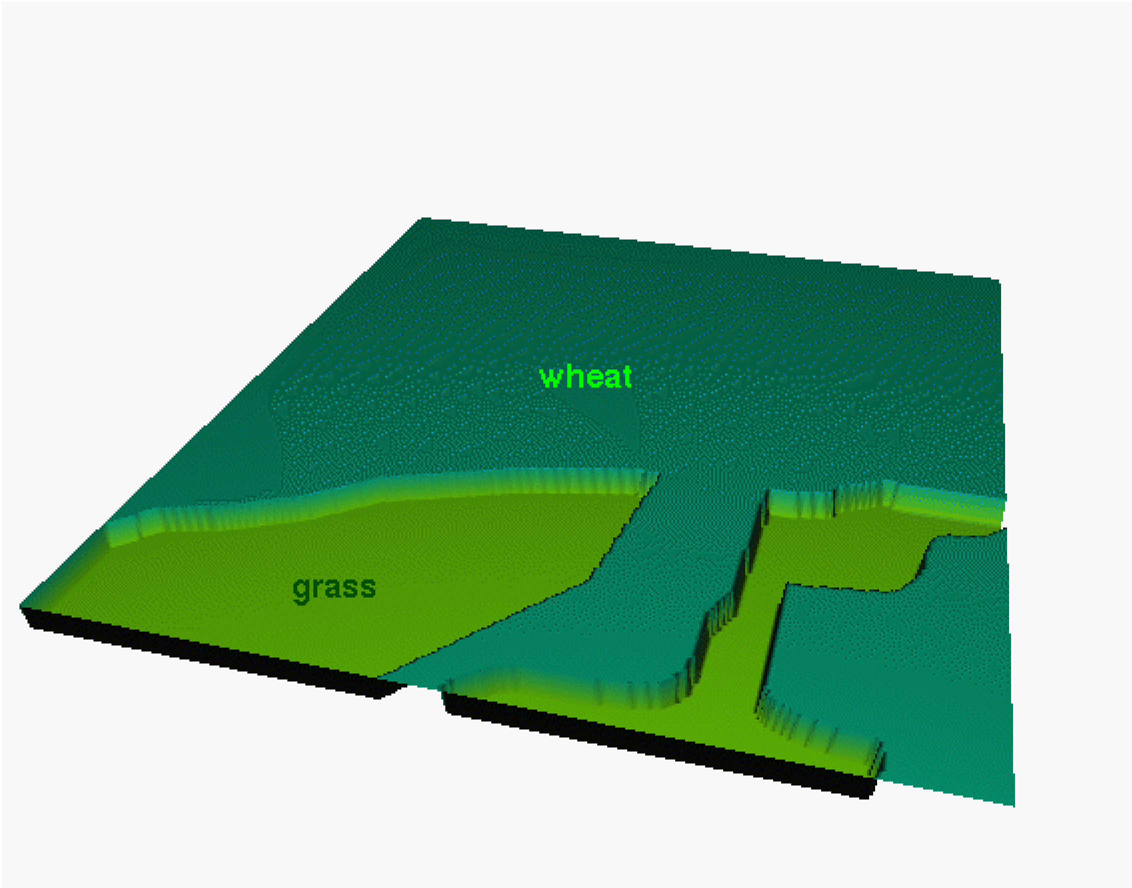
G Weintraub, A Jones, A Magendzo, M Meacham, and M Kirkby. A Heuristic System to Solve Mixed Integer Forest Planning Models. *Operations Research*, 42(6):1010--1023, 11 94.

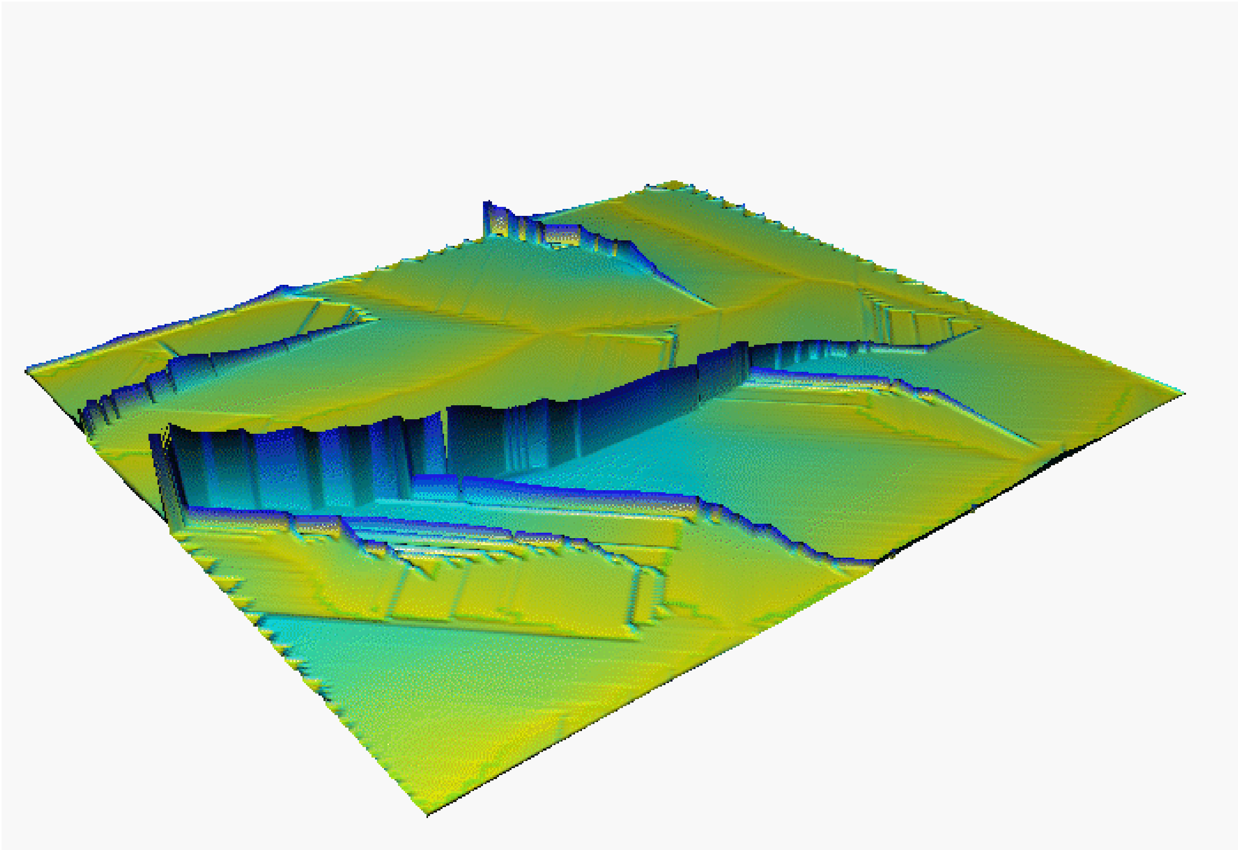
Up: Solving the Clear-Cut Scheduling **Previous:** 6 Acknowledgements

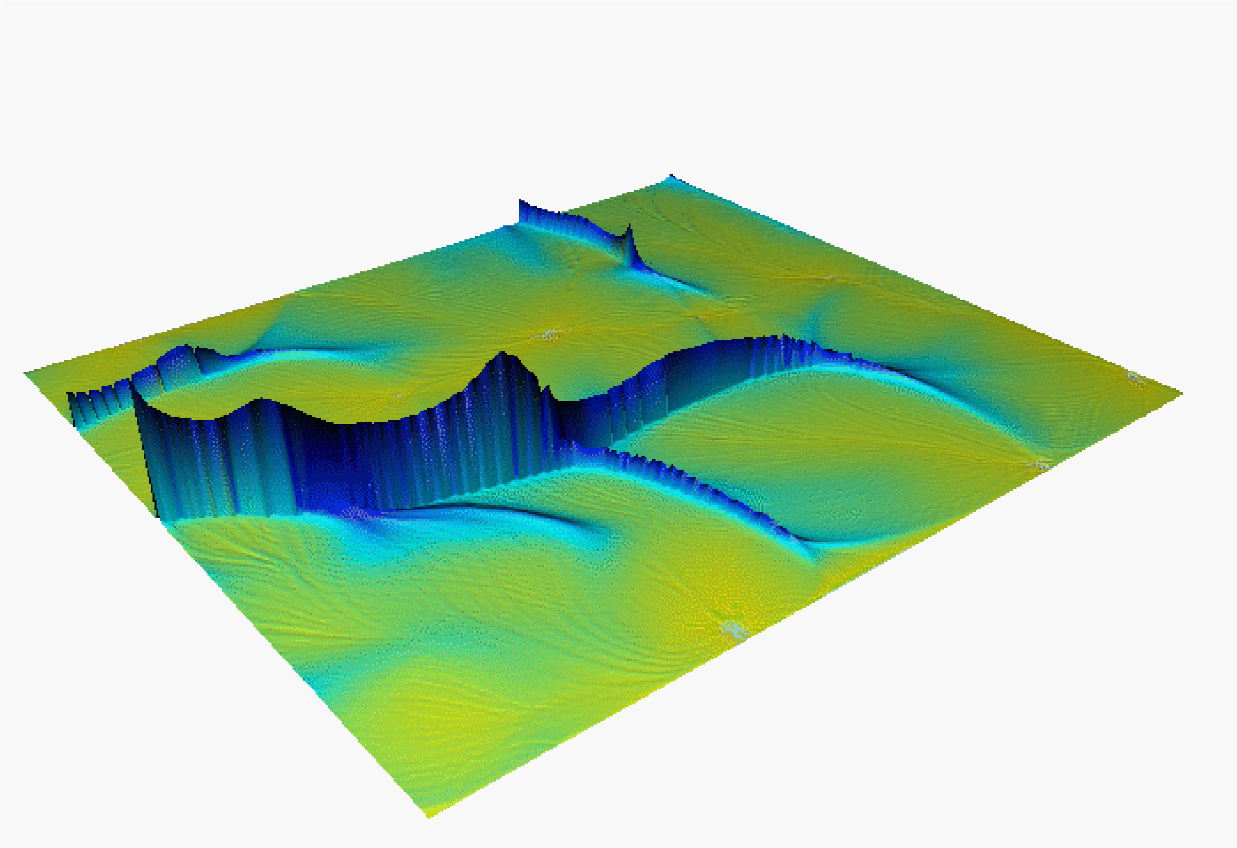
gunnar.misund@si.sintef.no

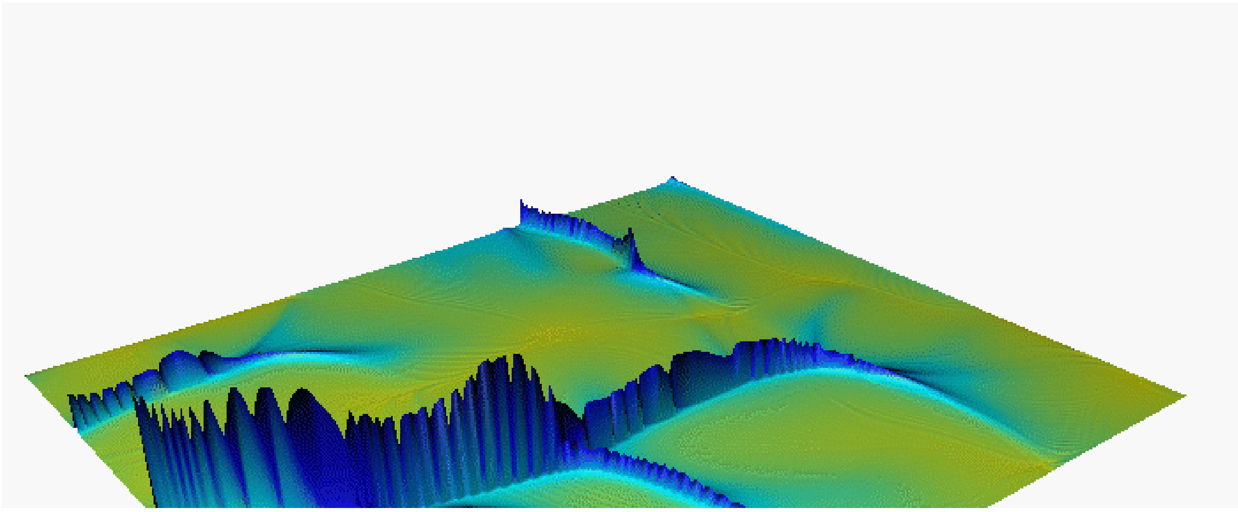

Elevations and data points
100 m

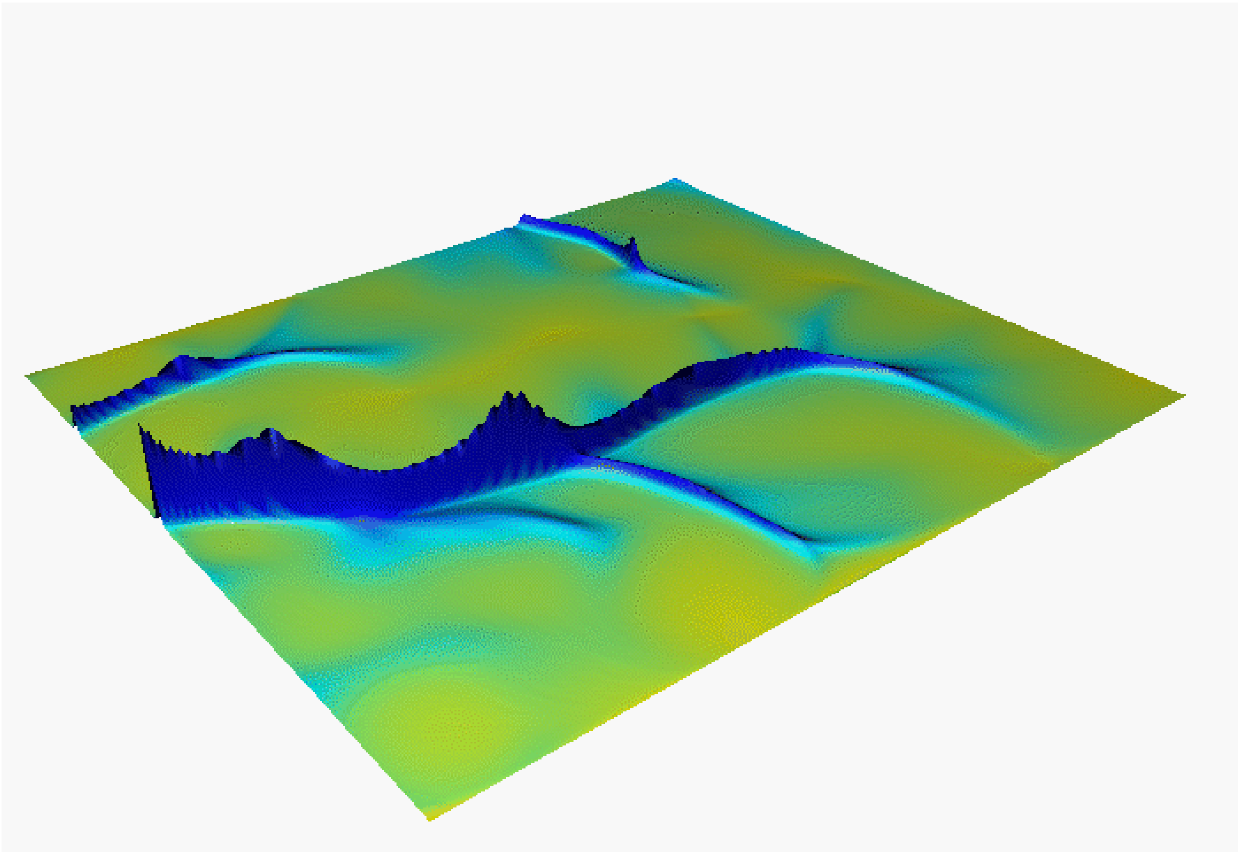

Aspect
90
180
270
360
100 m
Helena Mitasova 1993
Slope
100 m
Helena Mitasova 1993
Profile curvature
0) no data
1) −0.02 convex
2) −0.01
3) −0.005
4) −0.001
5) 0.
6) +0.001 concave
7) +0.005
8) +0.01
9) +0.02
10) +0.03
100 m
Helena Mitasova 1993
Tangential curvature
0) no data
1) −0.02 convex
2) −0.01
3) −0.005
4) −0.001
5) 0.
6) +0.001 concave
7) +0.005
8) +0.01
9) +0.02
10) +0.03
100 m
Helena Mitasova 1993

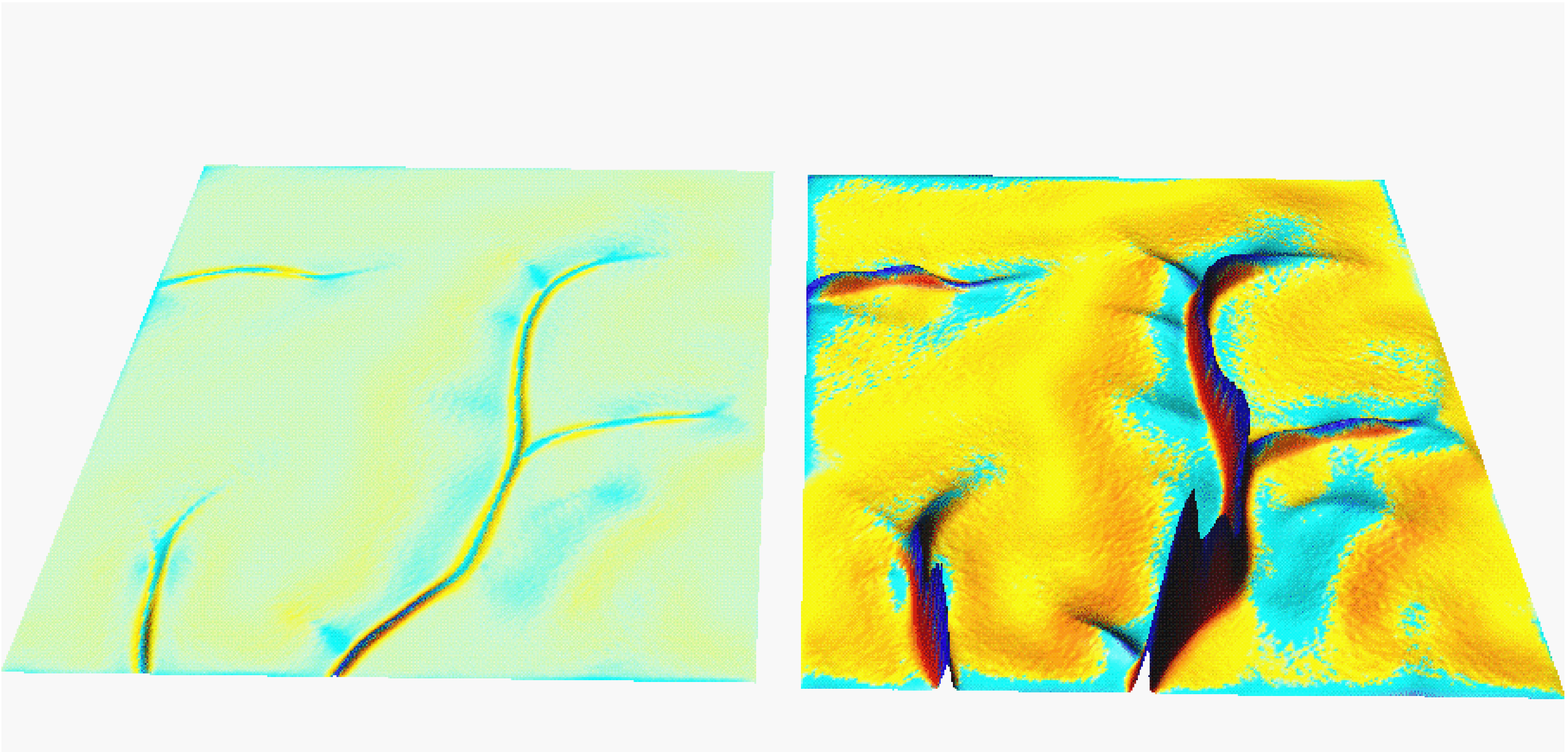

wheat
grass

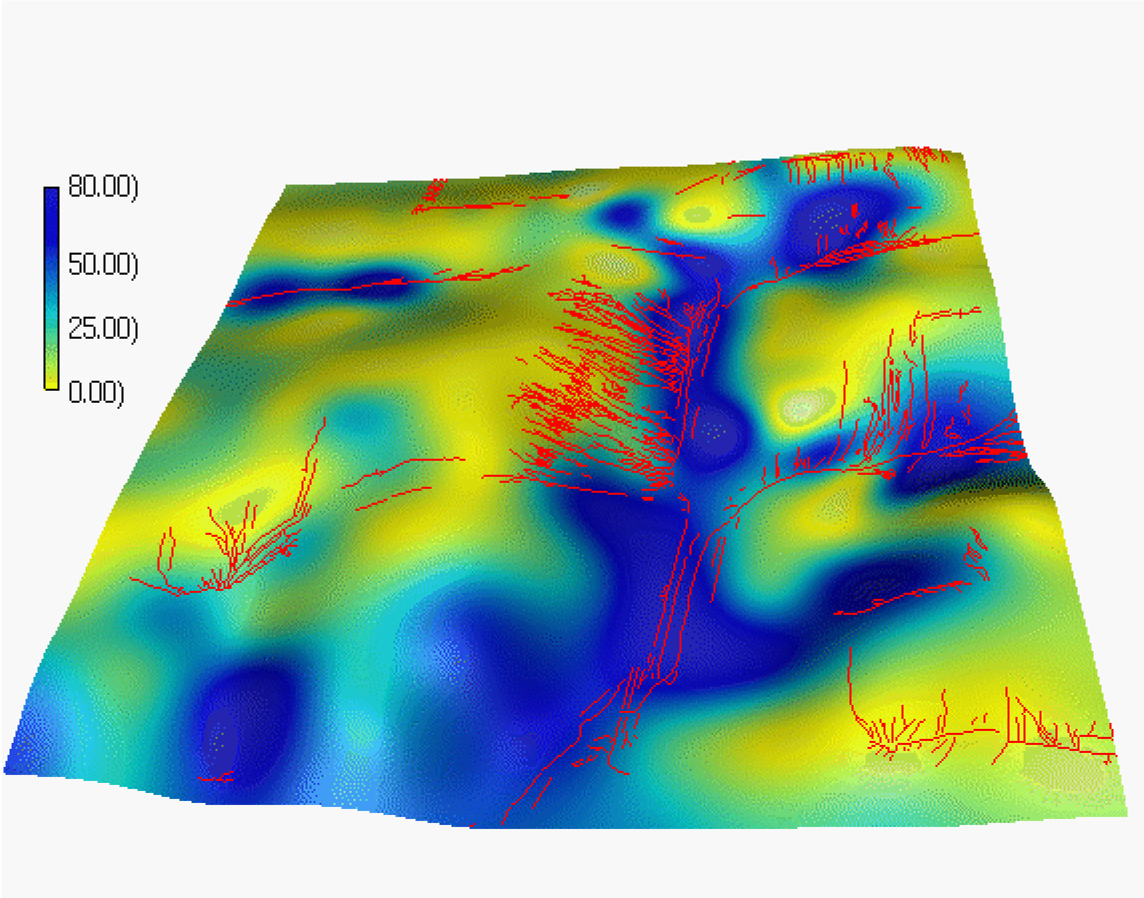


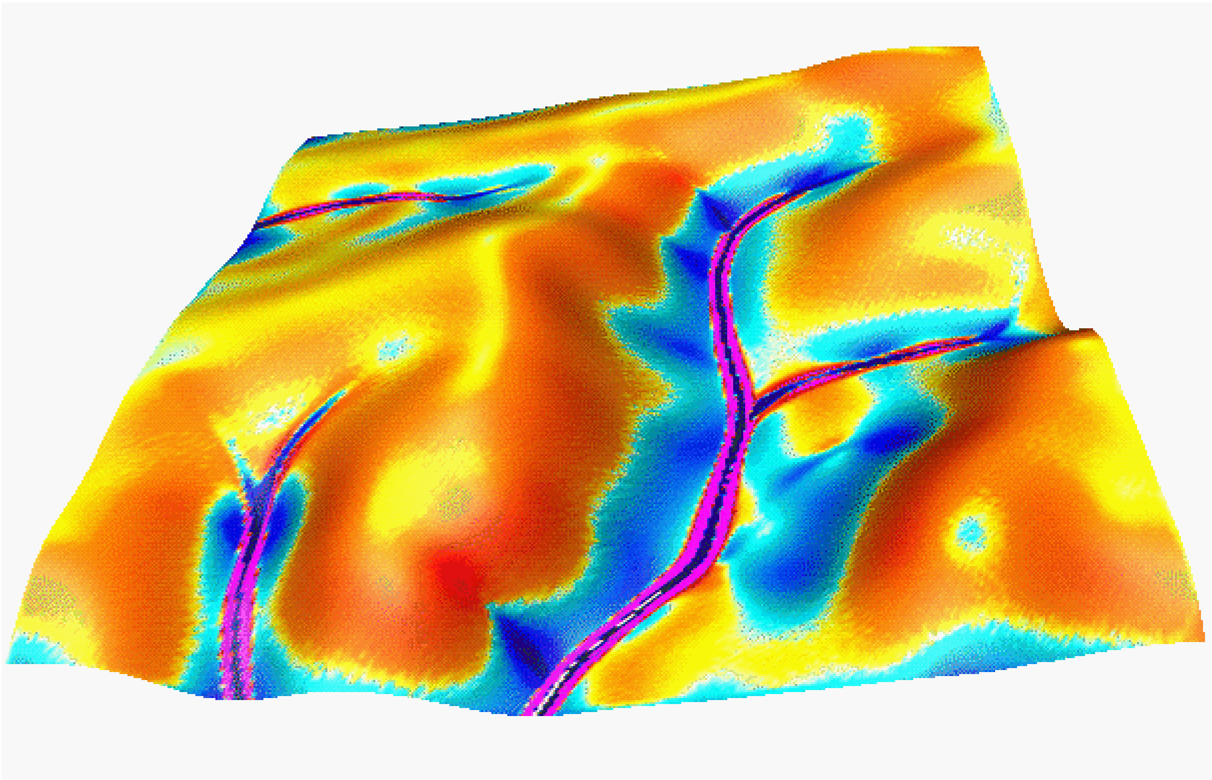


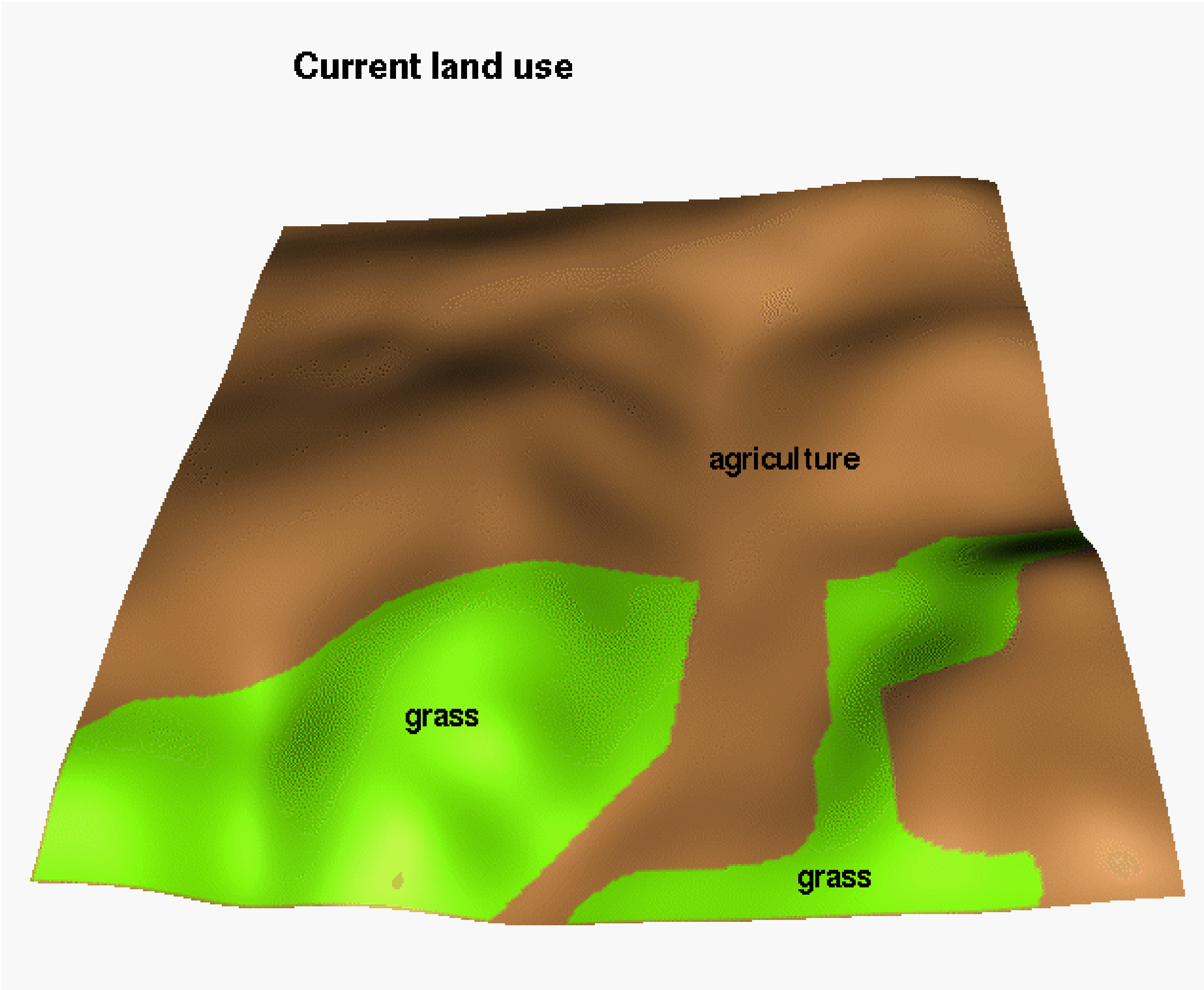


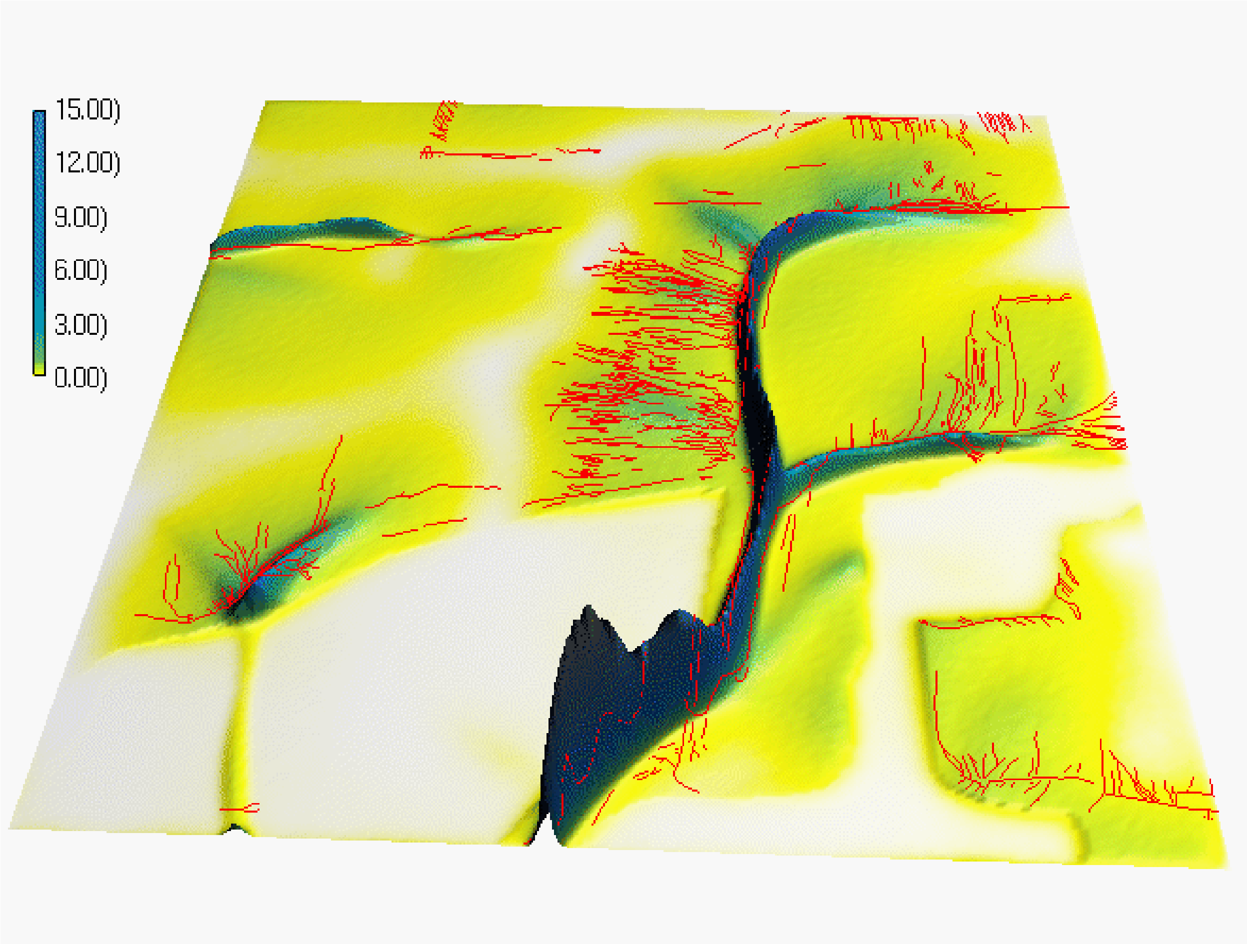


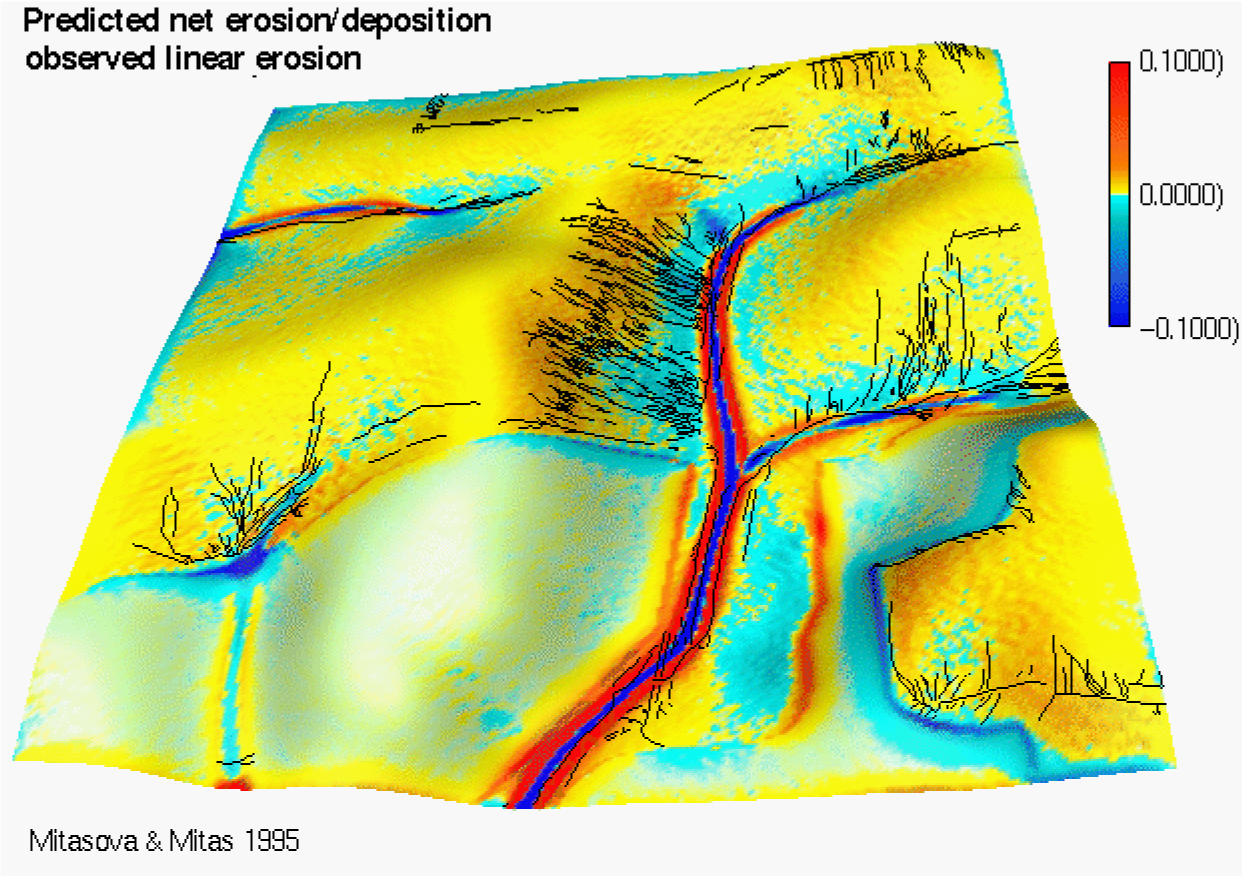


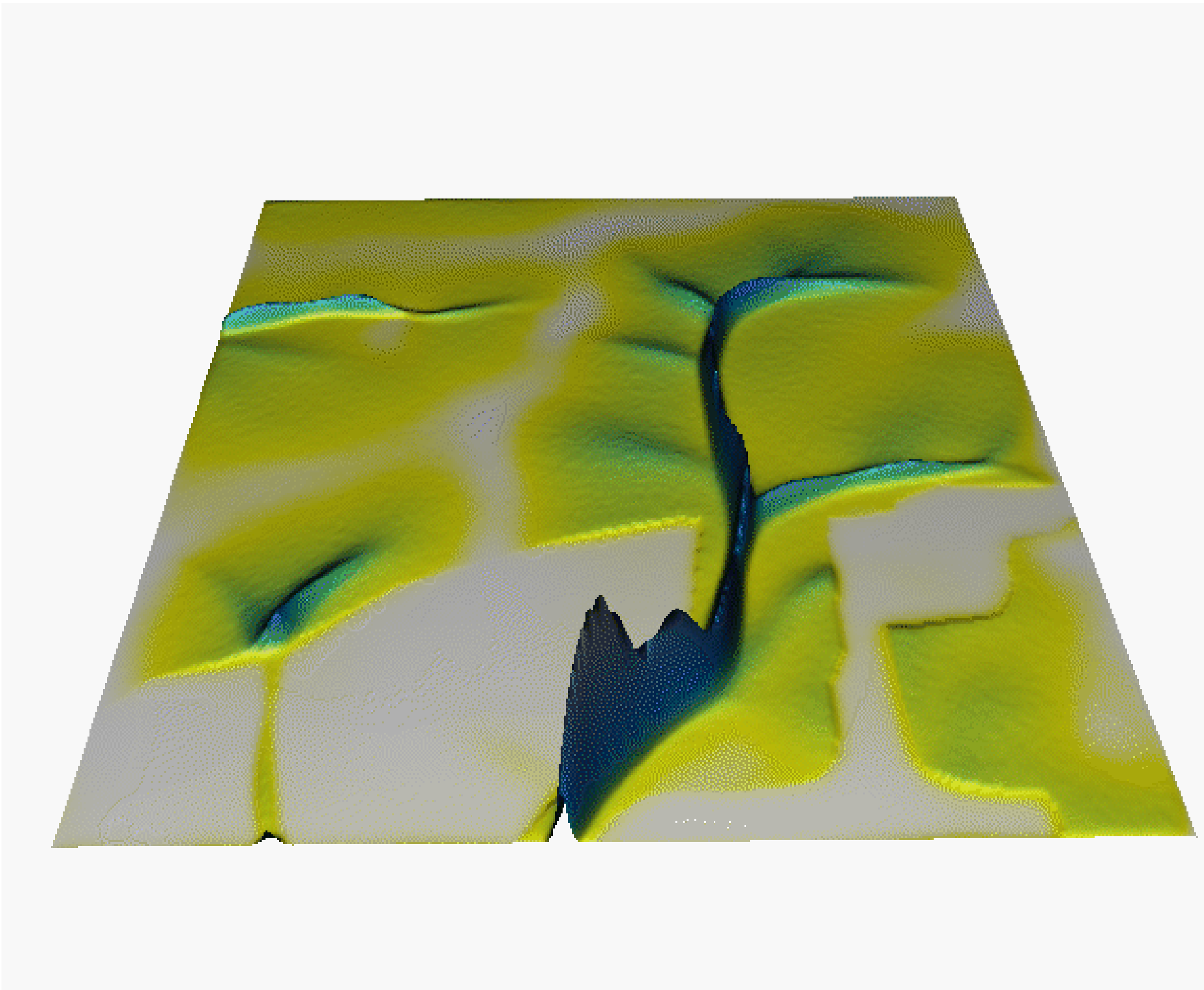

80.00)
50.00)
25.00)
0.00)

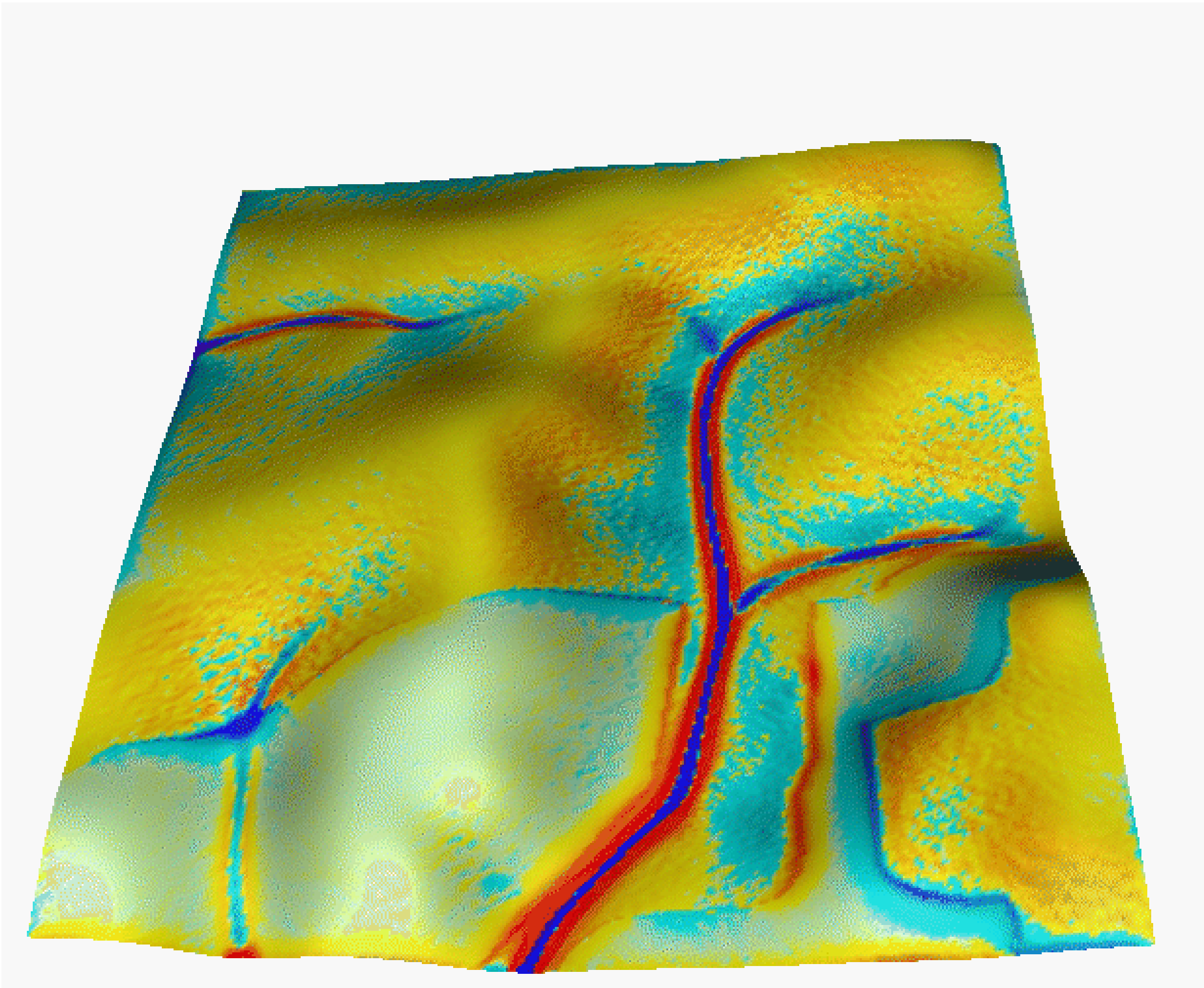


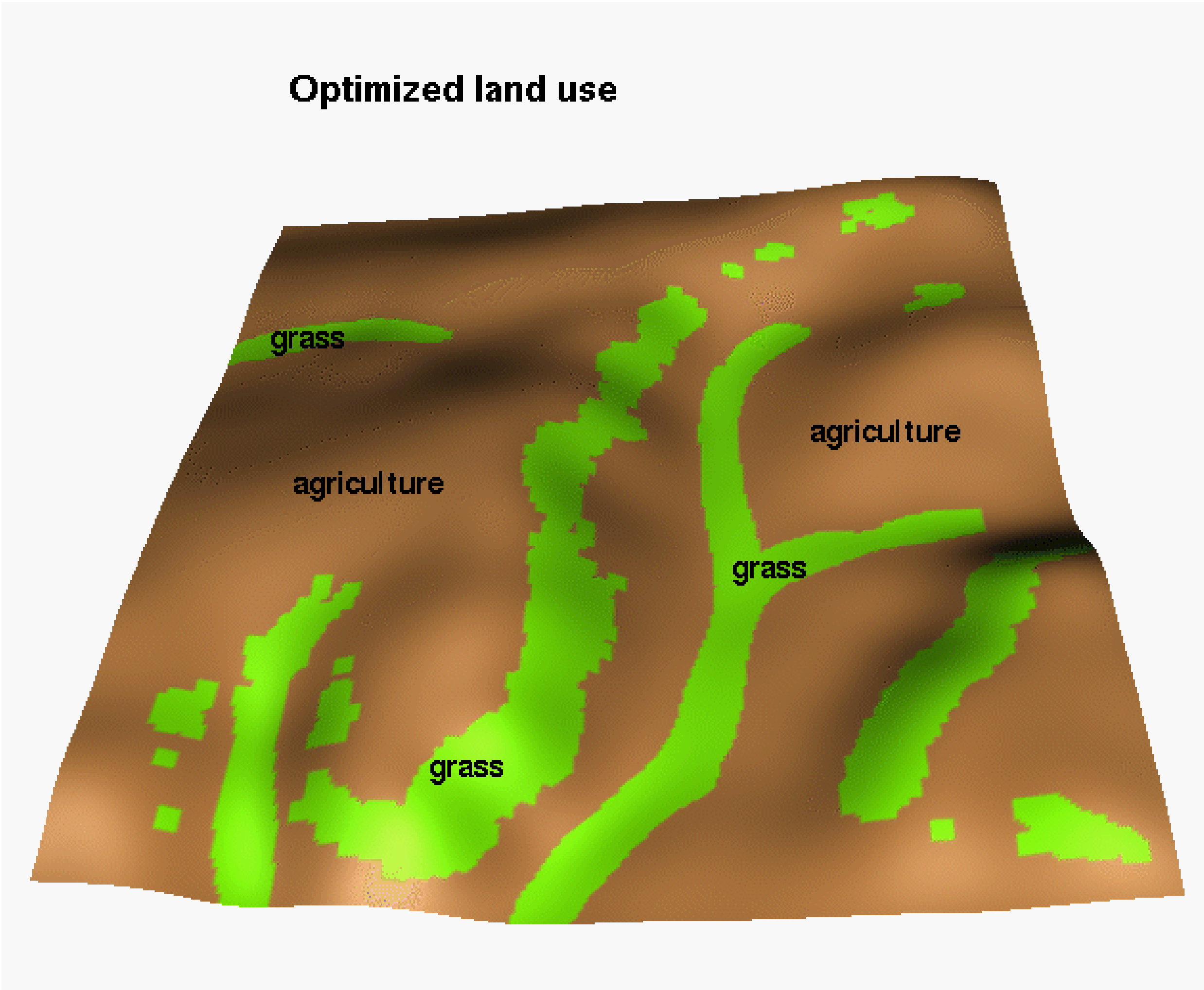

Current land use
agriculture
grass
grass

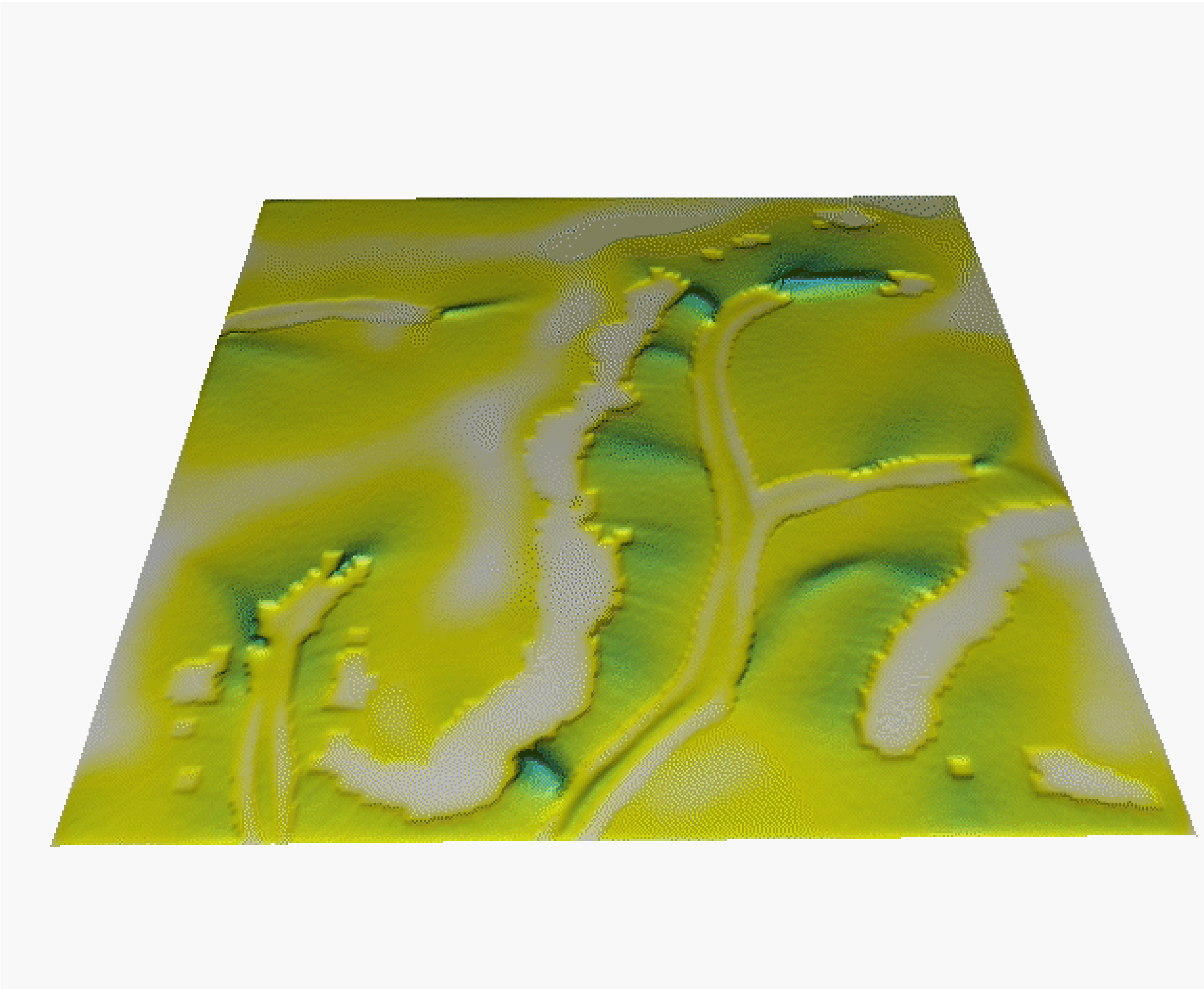

15.00)
12.00)
9.00)
6.00)
3.00)
0.00)

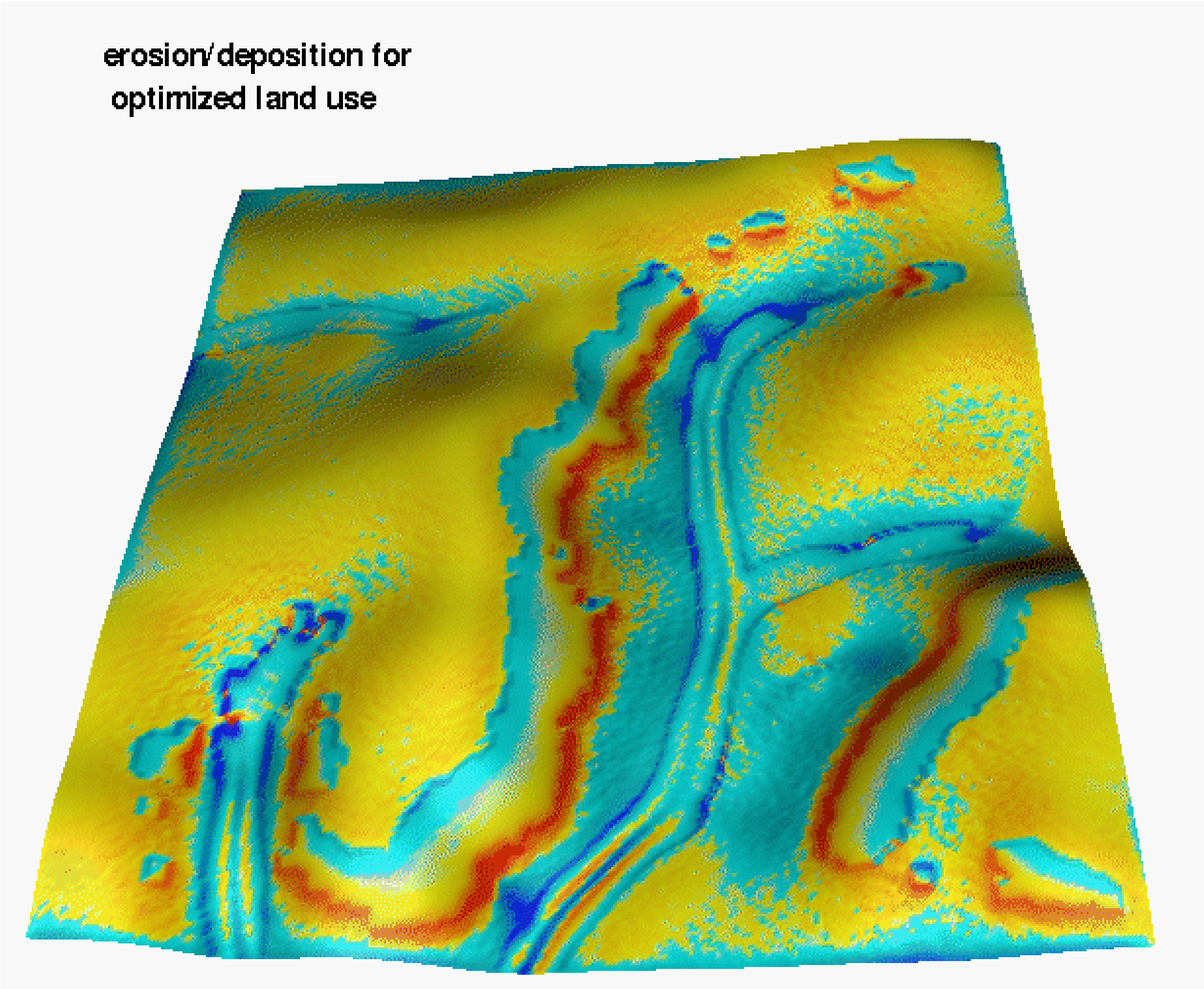

Predicted net erosion/deposition
observed linear erosion
0.1000)
0.0000)
−0.1000)
Mitasova & Mitas 1995

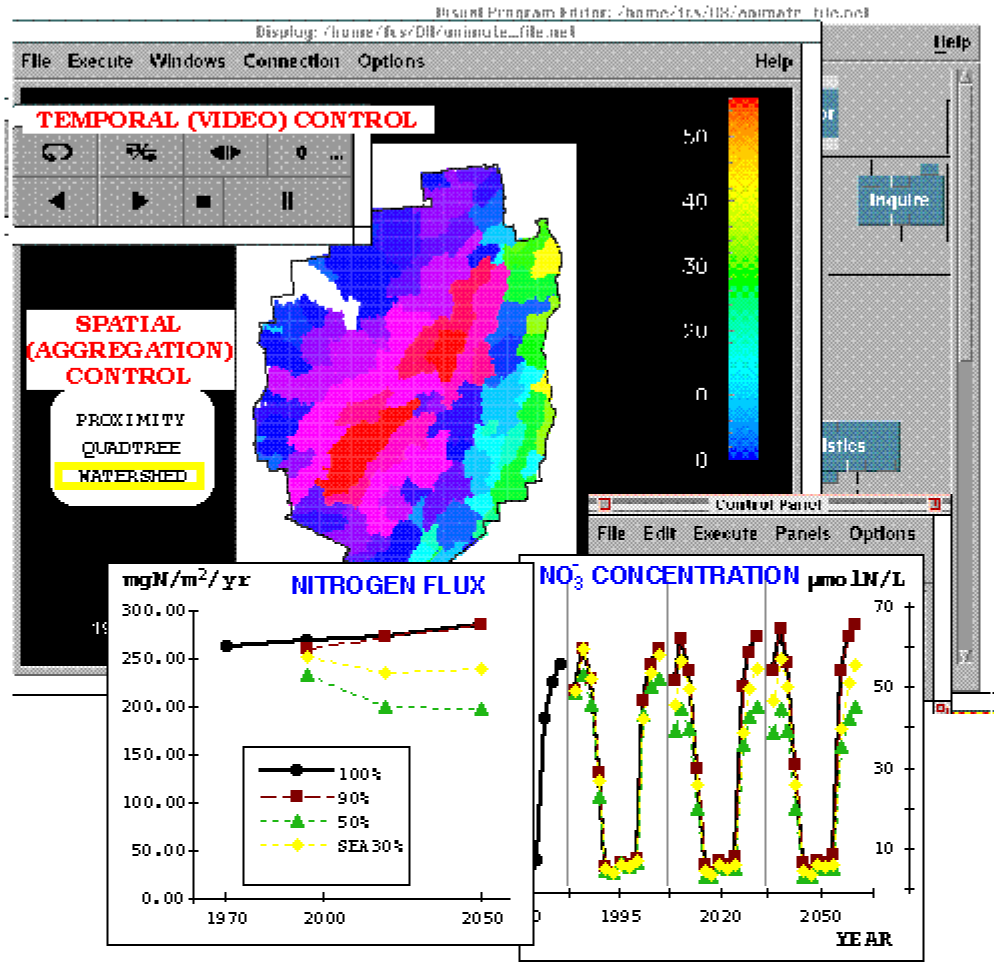





Optimized land use
grass
agriculture
agriculture
grass
grass




erosion/deposition for
optimized land use



File Execute Windows Connection Options Help
TEMPORAL (VIDEO) CONTROL
SPATIAL (AGGREGATION) CONTROL
PROXIMITY
QUADTREE
WATERSHED
Inquire
Control Panel
File Edit Execute Panels Options
mgN/m2/yr
NITROGEN FLUX
300.00
250.00
200.00
150.00
100.00
50.00
0.00
100%
90%
50%
SEA30%
1970
2000
2050
NO3 CONCENTRATION
µmol N/L
70
50
30
10
1995
2020
2050
YEAR

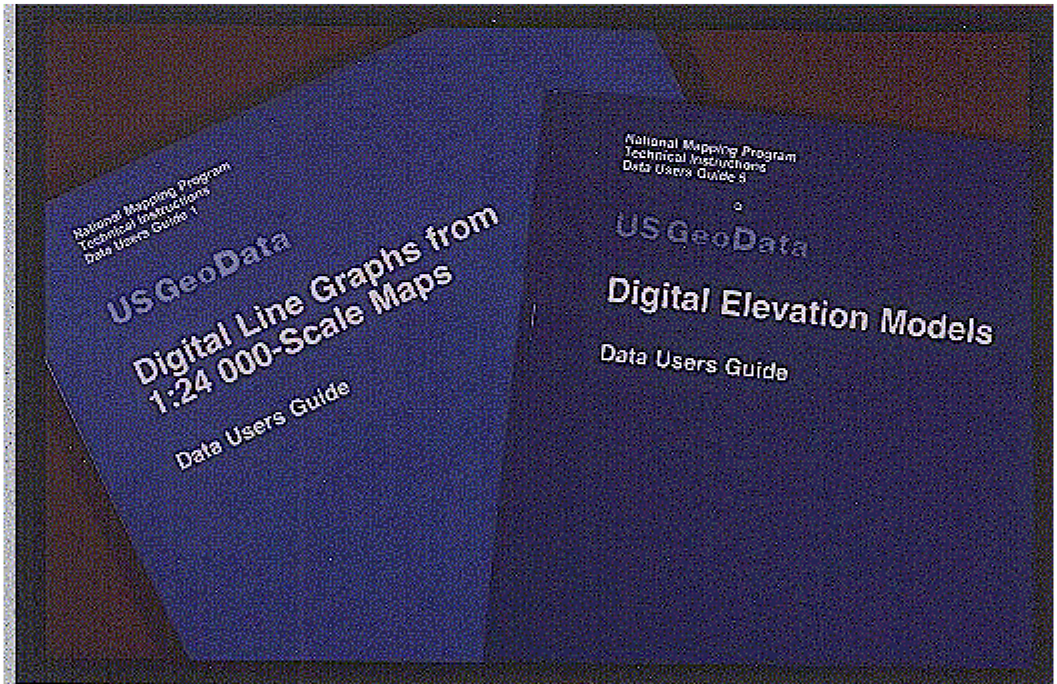


Figure 1.--National Mapping Program, USGeoData, Data Users Guides provide information about the Digital Elevation Data and Digital Line Graphs produced at the U.S. Geological Survey. These documents can be obtained from any USGS Earth Science Information Center or by calling 1-800-USA-Maps.

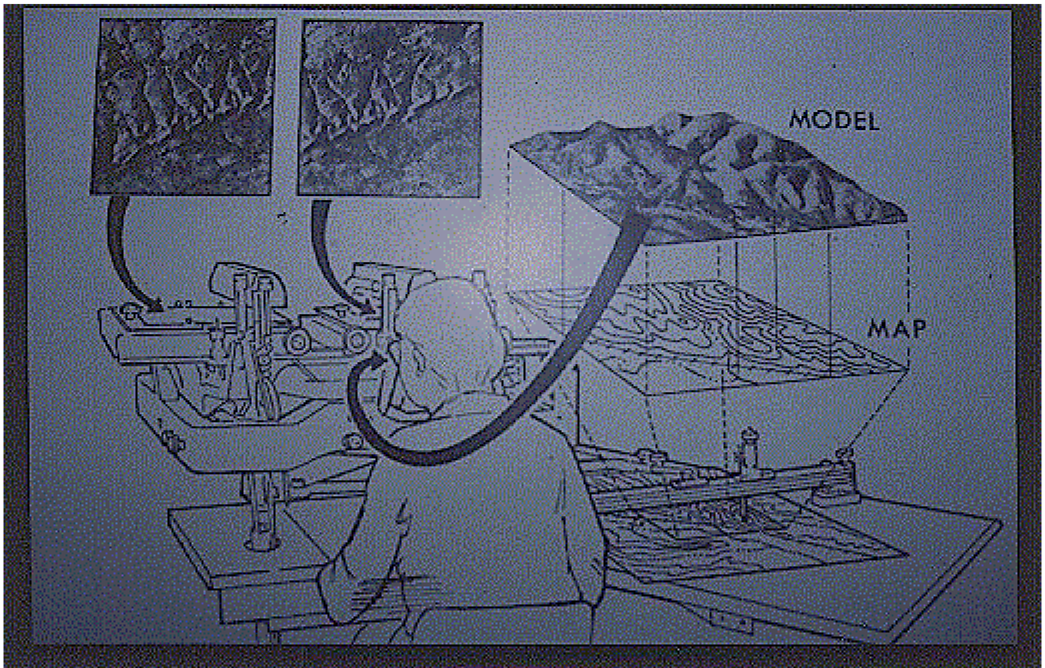


Figure 2.--Illustration of the photogrammetric compilation of a topographic map from stereo aerial photography. The Contours depicted on topographic map can be scanned and used in the production of a Digital Elevation Model.

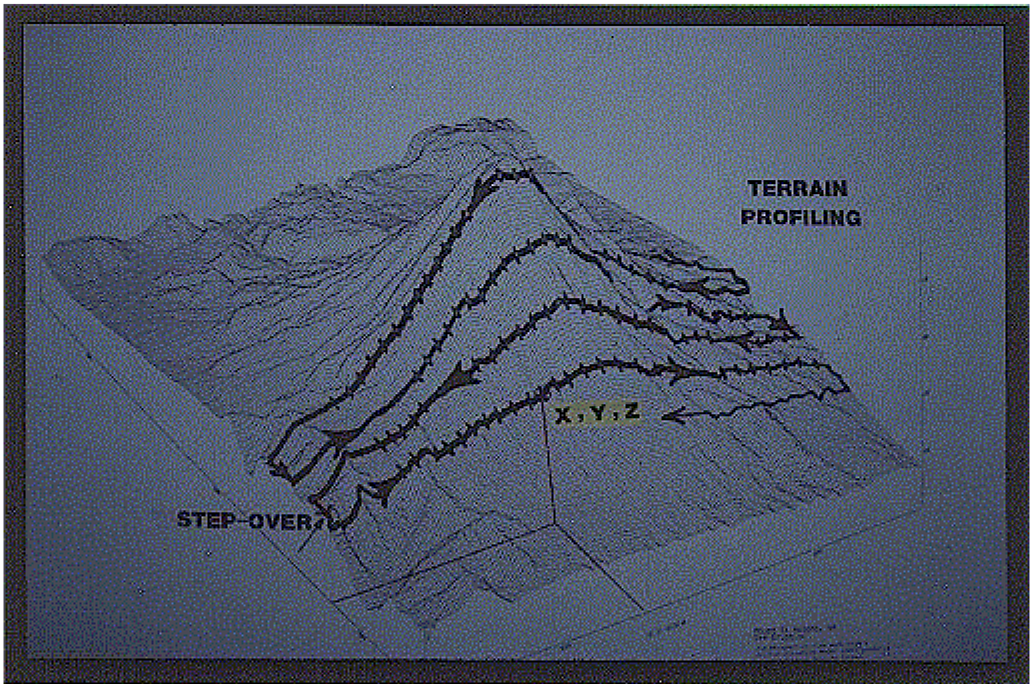


Figure 3.--Terrain Profiling is the collection of elevation data from a computer-assisted stereoplotting instrument, as shown in Figure 2. using digital encoders connected to a computer. The operator insures contact with the three-dimensional image of the land by keeping a floating dot centered in the field of view on the image surface.

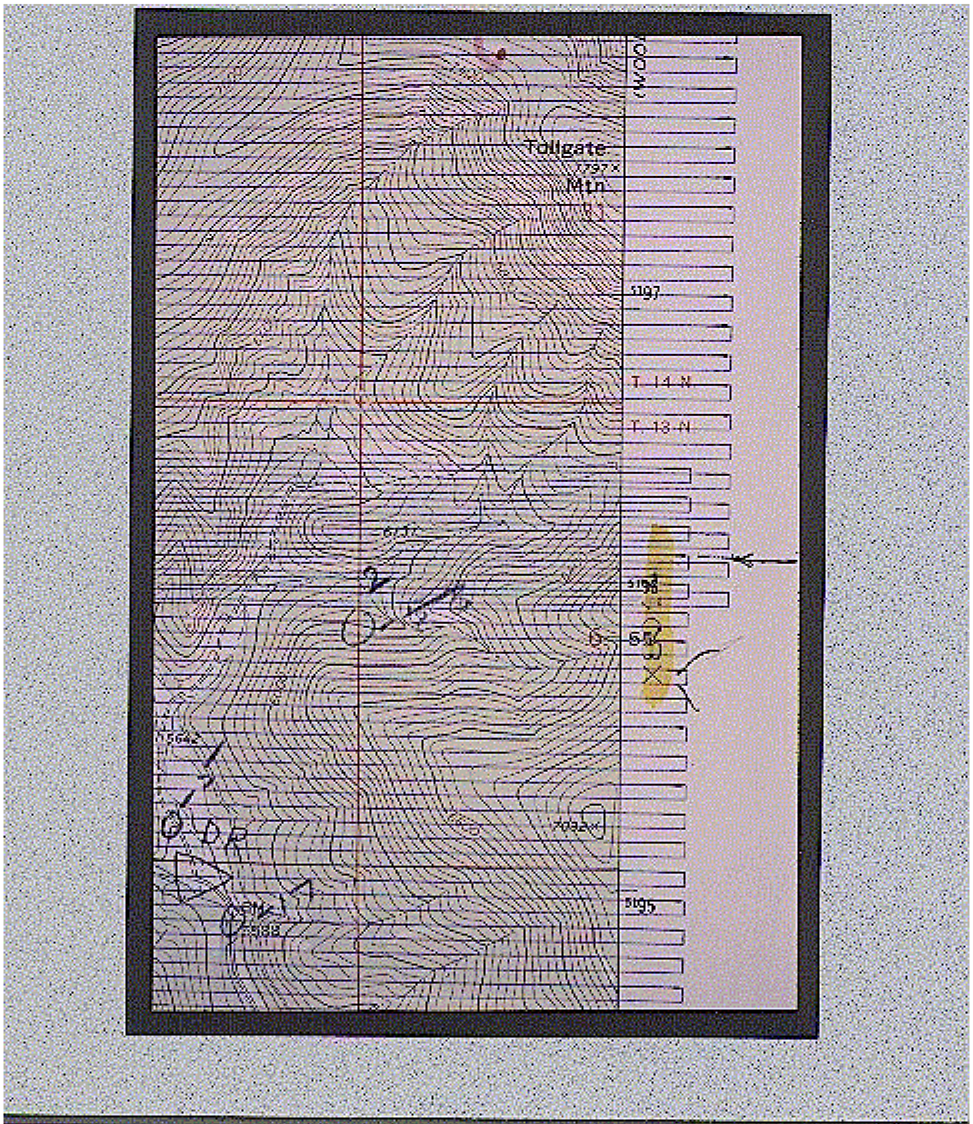


Figure 4.--A Graphic of the relationship of the tracks from the photogrammetric terrain profiling to the 1:24,000-scale 7.5-minute topographic map. Notice the profile overlap area between adjacent stereomodels.

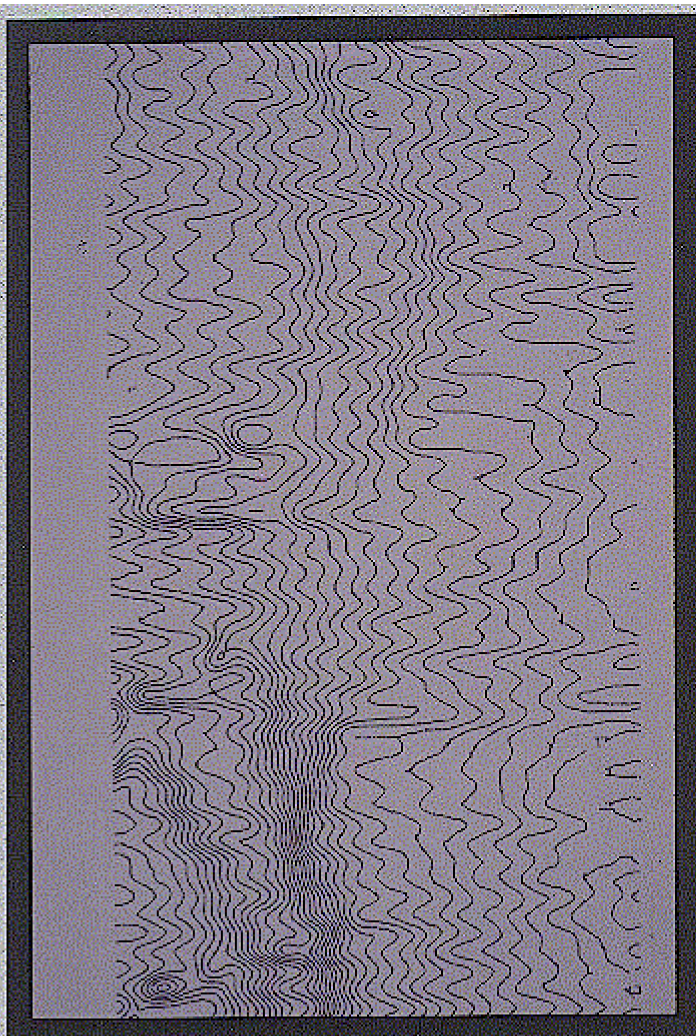


Figure 5.--A 20-foot contour plot illustrating the anomaly that may be detected from profiled data. The vertical displacement from the profile are usually 1-meter to 3-meters caused by the operator collecting slightly above the ground surface in one direction and slightly below the ground surface in the opposite direction of seperate profiles.

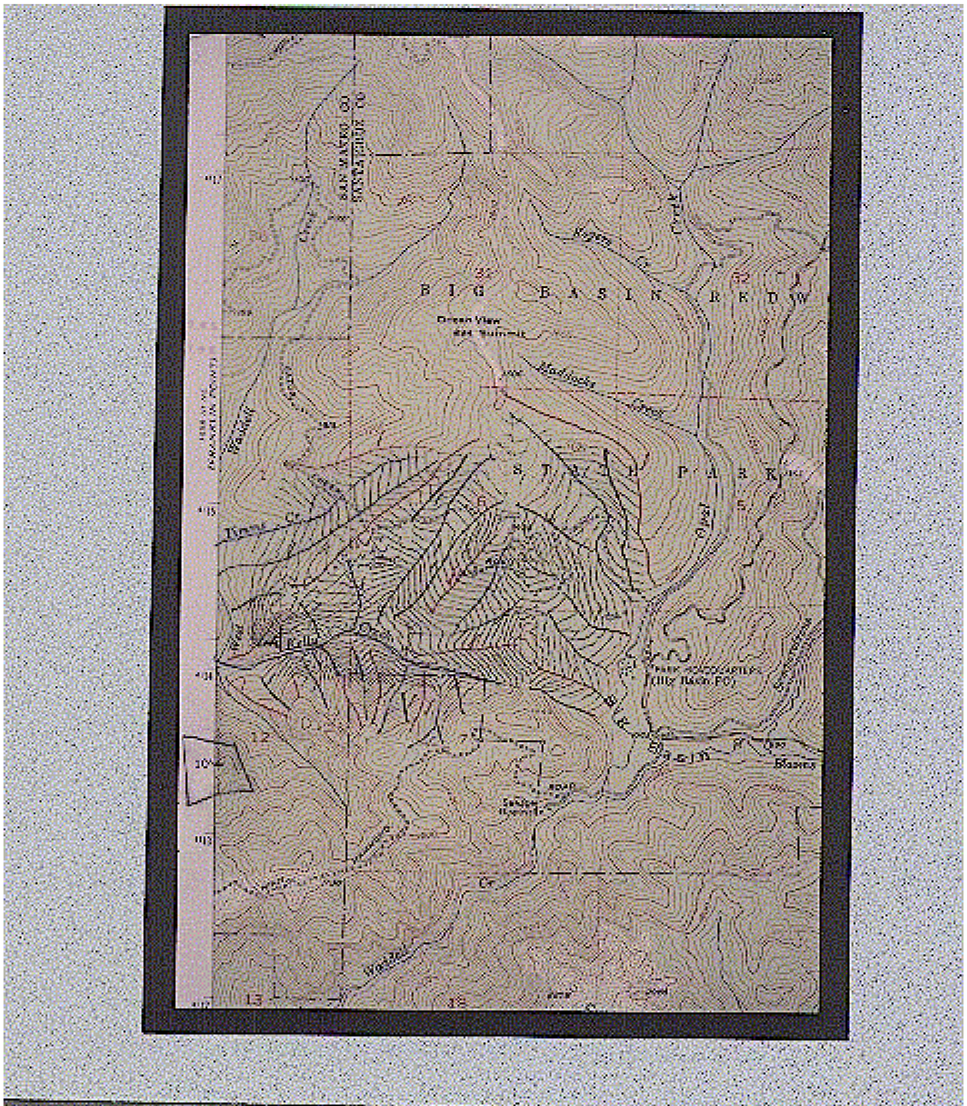


Figure 6.--Tracks from collecting attributed ridge lines (red), attributed drainage lines (blue), and attributed breaklines (heavy black). The thin black lines were plotted contour lines generated from a regular spaced elevation grid developed from a triangulated surface and compared to the contours from the 7.5-minute topographic map.

Figure 9.--Artifacts in elevation data may result from the type of gridding algorithm and the spatial location of the input x,y,z data. Smoothing can be applied the elevation model to give the aesthetic appearance of surface continuity, but in reality these elevation data may not reflect the accuracy of the original terrain surface.

Figure 8.--Edge joining between different elevation models is important when mosaicing 2 or more models. Some software packages will provide a blending algorithm to insure continuity of surface elevation data, while others will just butt the edges as they exist as shown.

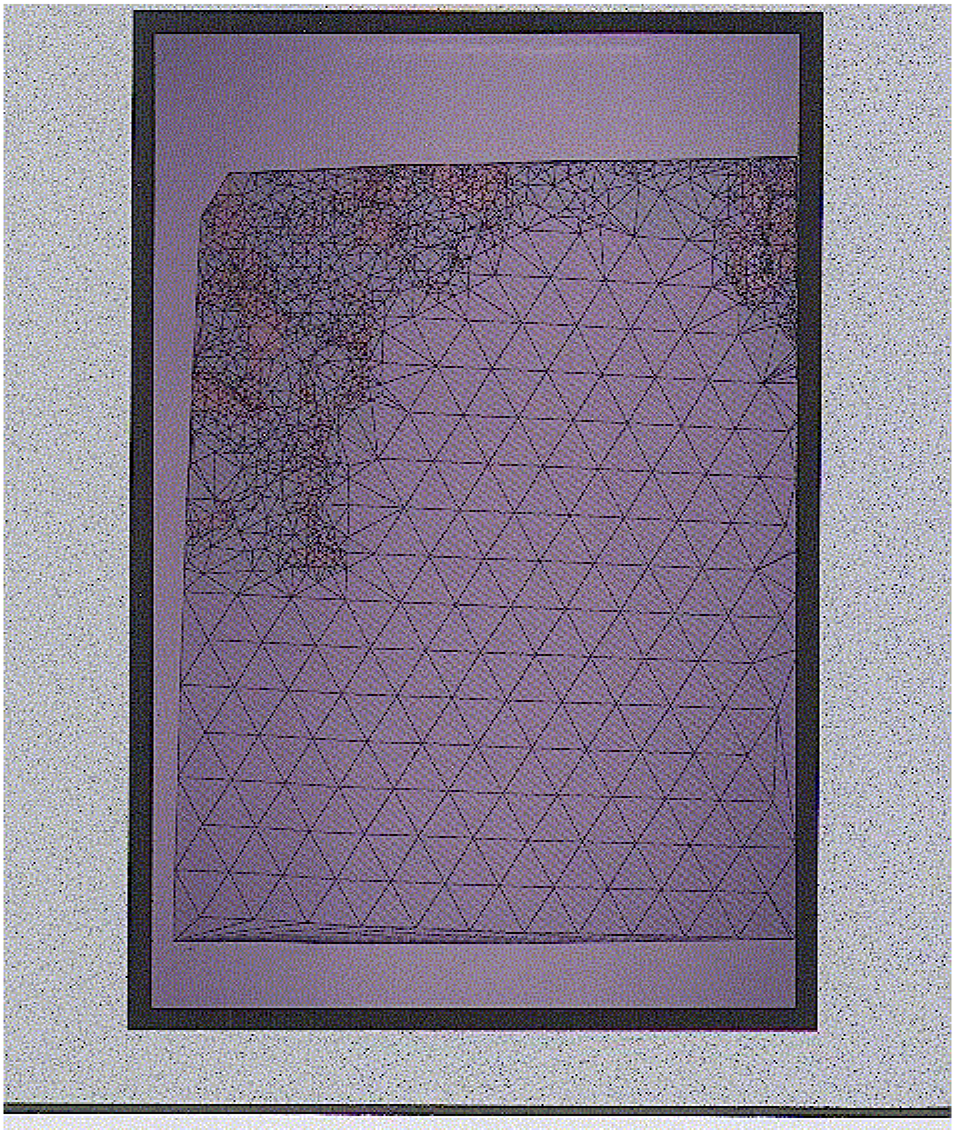


Figure 7.--A Triangulated Irregular Network plot generated from digital contour data vectorized a 7.5-minute topographic quadrangle hypsographic theme. The regular shaped trianlges are the surface of the water. Note that the shoreline and drainage patterns are not distinguishable as a result of a lack of a breakline information.

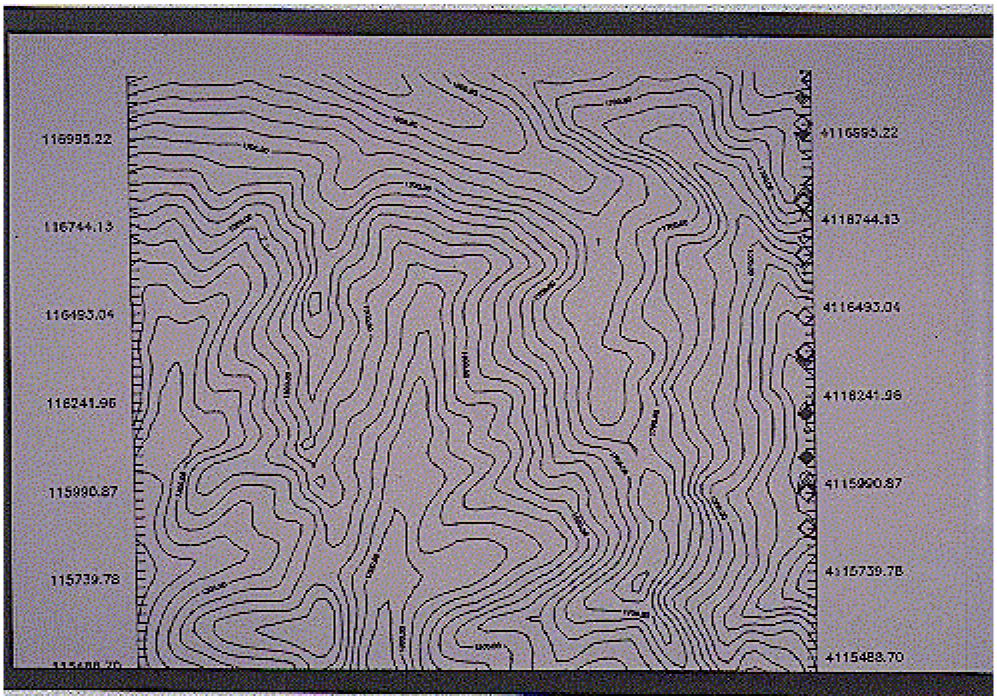


Figure 10.--The contour plot demonstrates that contours generated from regular gridded data or TIN data can produce artifacts possibly due to the original grid spacing, accuracy of the original grid, choice of points used in the TIN, and the contour interval chosen.

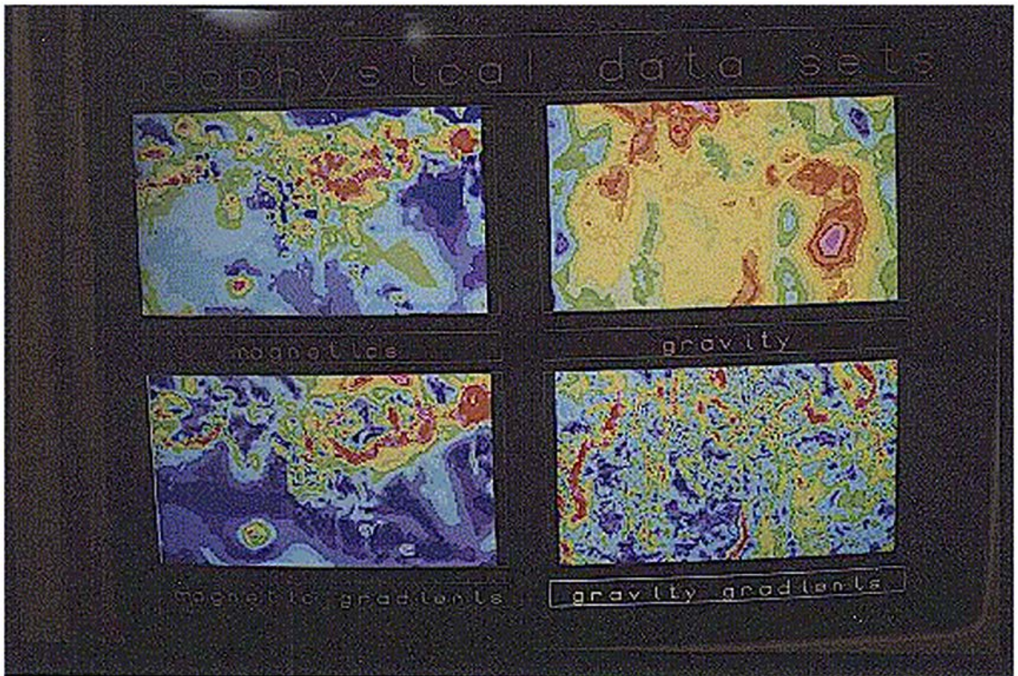


Figure 11.--Elevation data is used in geophysical applications to portray magnetics, magnetic gradients, gravity, and gravity gradients.

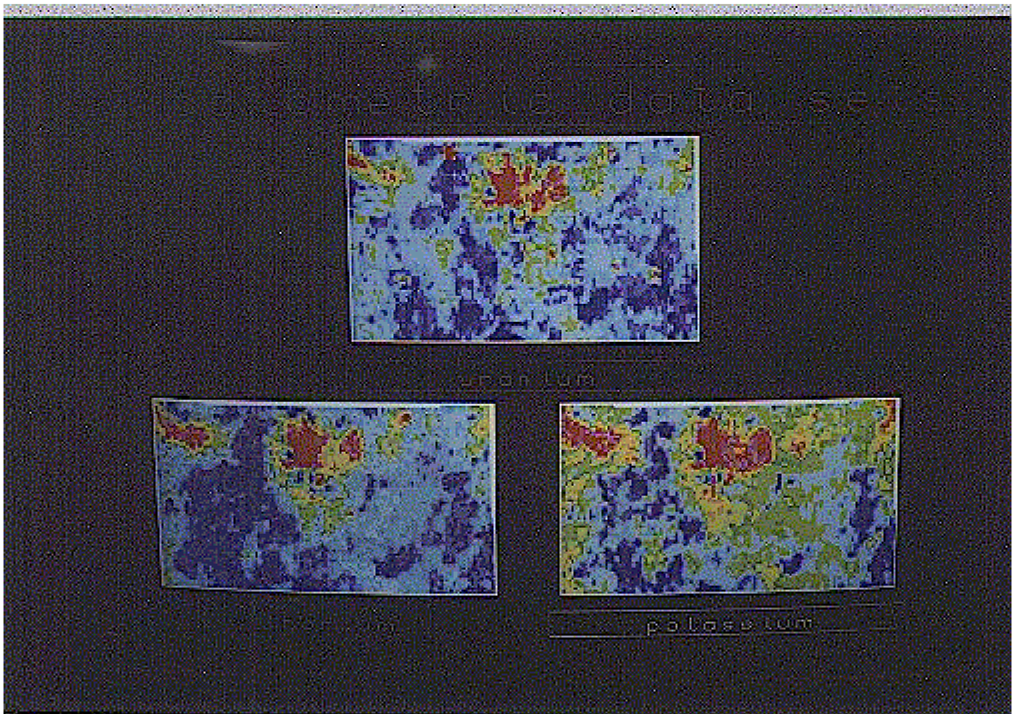


Figure 12.--Elevation data is useful in aiding in the exploration of mineral deposits.

Figure 13.--Example of digital shading overlaid on a 7.5-minute topographic quadrangle. This capability can also used with a scanned Digital Raster Graphic, thereby creating the shading and graphic all in digital form.

Up: Application of CAD Frameworks **Previous:** Application of CAD Frameworks **Next:** Figure 2: NELSIS

Figure 1: HYBROW

The HYBROW (Hydrologic Browser) user interface for a collaborative research project involving 4 experimental catchment sites in Belgium, France, and Italy.

A typical user sequence is as follows:

1. Select the catchment site of interest.
2. Select a data type from the pop up menu that appears by clicking on the site icon.
3. Push the ``populate data objects'' button to get the symbolic representations to appear on the screen.
4. Select one of the objects, for instance a digital elevation model (DEM) of the catchment site.
5. Select the desired function from the pop up menu, for instance a visualization module.

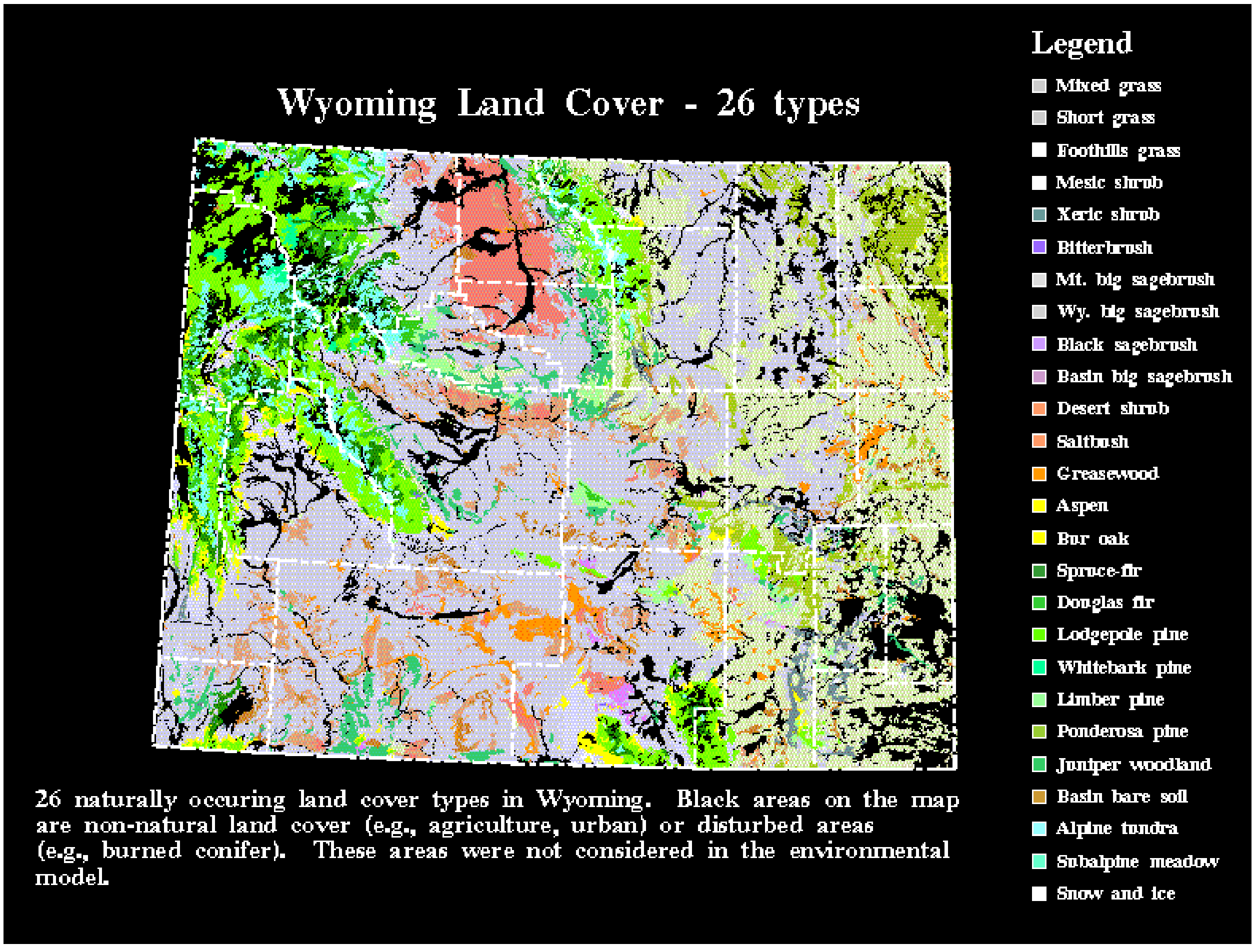
Up: Application of CAD Frameworks **Previous:** Figure 1: HYBROW

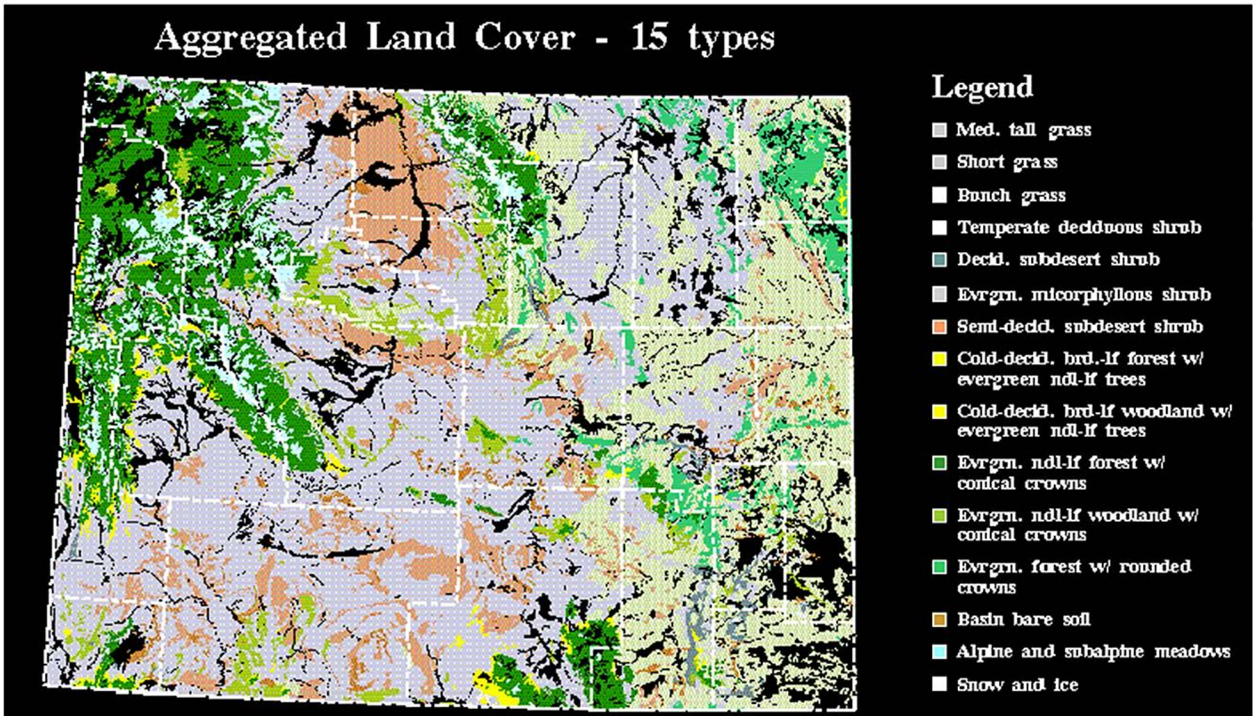
Figure 2: NELSIS

The NELSIS Design System User Interface applied to a hydrologic modeling process.

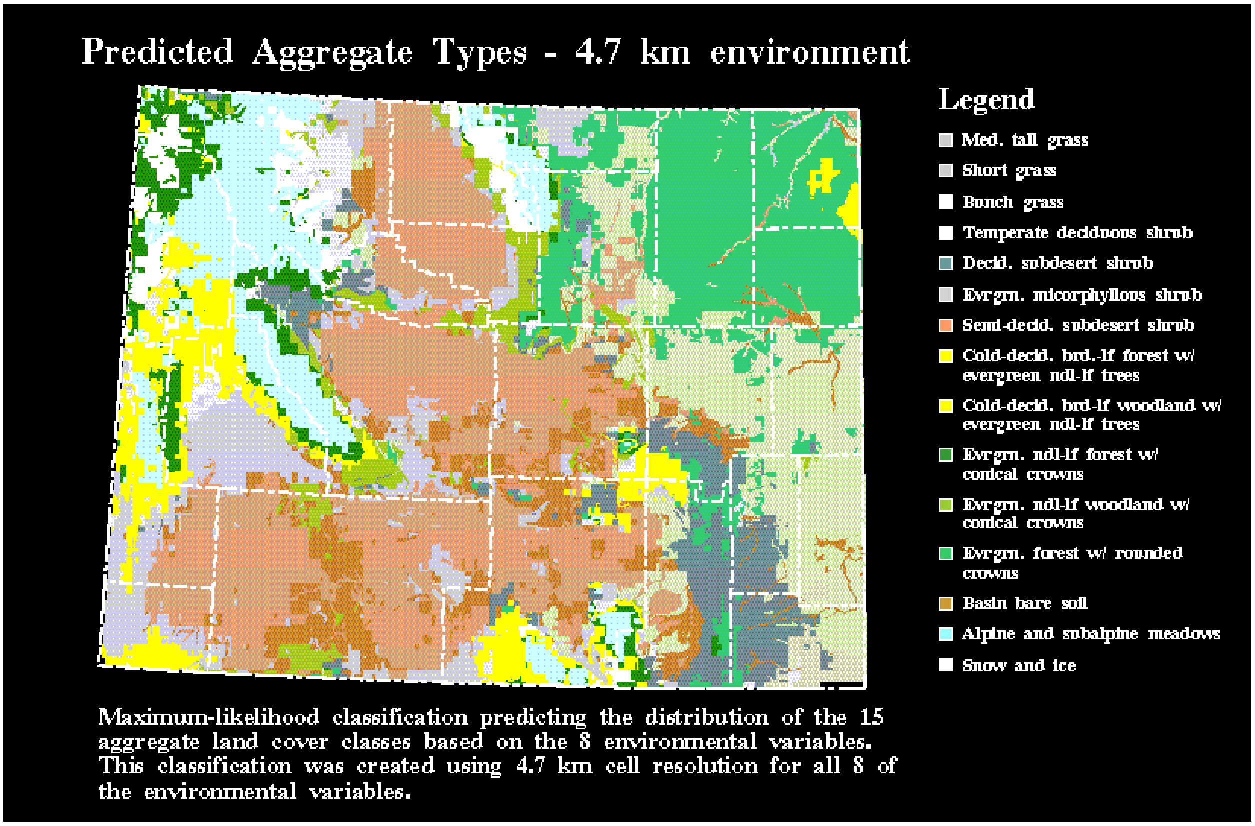
The design flow shown in the image was developed using the NELSIS FLow User Interface Designer (FLUID). The design flow for the hydrologic application is divided into 3 parts:

1. Input of the model parameters and data, including selection of catchment and rainstorm events to be simulated. The inputs can be edited and versioned as they are selected.
2. Model execution (this can only be done when all inputs are available; data availability is indicated by red/green coloring).
3. Viewing of intermediate simulation results (*view_monitor*) and analysis of the results (using visualization systems such as AVS and Iris Explorer).


Wyoming Land Cover - 26 types
Legend
Mixed grass
Short grass
Foothills grass
Mesic shrub
Xeric shrub
Bitterbrush
Mt. big sagebrush
Wy. big sagebrush
Black sagebrush
Basin big sagebrush
Desert shrub
Saltbush
Greasewood
Aspen
Bur oak
Spruce-fir
Douglas fir
Lodgepole pine
Whitebark pine
Limber pine
Ponderosa pine
Juniper woodland
Basin bare soil
Alpine tundra
Subalpine meadow
Snow and ice
26 naturally occuring land cover types in Wyoming. Black areas on the map are non-natural land cover (e.g., agriculture, urban) or disturbed areas (e.g., burned conifer). These areas were not considered in the environmental model.



Map of 15 aggregated cover types in Wyoming. These types were created by merging the 26 Gap cover types into more broadly defined types based on physiognomy (e.g., evergreen needleleaf forest with rounded crowns) and corresponding to the UNESCO formation level.



Maximum-likelihood classification predicting the distribution of the 15 aggregate land cover classes based on the 8 environmental variables. This classification was created using 4.7 km cell resolution for all 8 of the environmental variables.

Exploratory Visualization of Environmental Data

Eva-Maria Stephan
Department of Geography - GIS
University of Zurich, Switzerland

environmental data

Environmental data represent *natural phenomena*, which are measured (sampled) on the earth surface in some spatial and temporal resolution, e.g., elevation data, vegetation, temperature, wind, sunshine duration.

Data are *sampled* at irregular or regular points, and s.t. preprocessed, e.g., interpolated. They are stored as large index arrays and analyzed computationally, e.g., by GIS or supercomputers.

Understanding and modeling of natural phenomena and their interrelationships require large volumes of *high quality* data and the *combination of many different* data sets.

environmental modeling and analysis

Environmental sciences integrate data from various sources, and derive new data upon complex *numerical modeling and analysis*. The results of such analyses often are very large in size and incomprehensible.

Furthermore, modeling and prediction introduce *uncertainty*, which cannot be disregarded in model creation, analysis and decision making.

To address this problem, scientists have turned to visualization, for investigation at intermediate stages of analysis or for post-processing.

This work focuses on developing a visualization environment which optimizes environmental modeling and makes spatial data analysis more efficient.

scientific data visualization

Scientific data visualization, i.e., the *display of behavior*, is used in many areas to gain insight into complex data. *Cartographic visualization* for example - the oldest science for spatial data visualization - defines various rules for presenting spatial data in maps, e.g., color coding, symbology.

Additionally, computer graphics supports advanced techniques, like surface / volume rendering, and animation for presenting multi-dimensional data in images. *Multi-dimensional dynamic data visualization* offers flexible ways for presenting and analyzing complex data, and verifying modeling results if graphics systems are designed carefully along with data characteristics and the needs of exploratory data analysis (EDA).

exploratory data visualization

To enable scientists to maximize insight into data, a higher man-machine communication bandwidth is desirable. The analyst should be able to interact with displayed information and to *look behind* images.

Exploratory Data Visualisation as shown and perceived in the examples in this project therefore requires:

- *powerful graphical representations* for multivariate spatio-temporal data
- embedded in an *interactive and intuitive environment*.

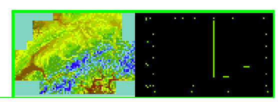
The proposed visualization environment combines realistic and analytic graphics. In addition it allows interactivity to control dynamically the visualization content, format, size and color and to query information directly from images.

data objects

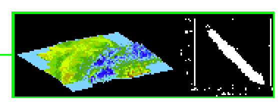
multi-dimensional data visualization*

**realistic representations of data keep spatial dimensions in the displayed image; analytic / statistic graphics disregard 'real-world' dimensions.*

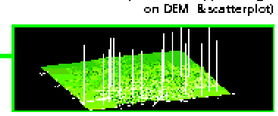
basic object types	realistic graphics	analytic graphics
2d spatial object	2d image	1d plot
2d temporal object	e.g., animated thermometer	1d curve
3d spatial object	3d color-coded image	2d plot
4d spatio-temporal object	animated 3d color-c. image	3d plot
data uncertainty	transparency, blinking, (fading) colors	probability surface



temperature image & histogram



DEM & temperature (mapped image on DEM & scatterplot)



location of original temperature sampling points displayed in DEM surface

This work proposes a new visualization environment for exploring visually characteristics, uncertainties and correlations of spatial data objects, it is based on the integration of realistic and analytic visualization techniques and their interactive exploration, i.e., the interactive manipulation of presentation parameters and the interactive selection of presented objects and visualization techniques.

January, 1996

DataScaping - Interactively Exploring Environmental Data

Eva-Maria Stephan
Department of Geography - GIS
University of Zurich, Switzerland

visualization techniques

DataScaping** is a visualization environment for exploring multivariate spatial and temporal data objects with various multidimen- sional interactive visualization techniques, such as:

- *analytic* graphics, e.g., time-curves, scatterplots, histograms, profiles etc.,
- *realistic* graphics, e.g., 2d color images, 3d wire- and rendered surfaces, 3d contours etc.,
- *pseudo-realistic* graphics, e.g., color mapping on attribute surfaces, animation of attribute data, etc., and
- *the combination of analytic and realistic* graphics by dynamic linking of windows.

The possibility to choose and compare manifold representations for spatial data help to explore data and to find the best view for each problem.

** DataScaping presents any data object as rendered surface, not only *landscapes*.

interactive analysis with viewers

Each visualization technique in DataScaping*** is implemented as an *independent module* and realized in a separate window, called a *viewer. Each* viewer provides control-buttons and sliders for interaction with presentation parameters and the object.

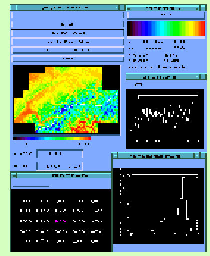
In a main viewer, data objects and visualization techniques may be selected from flexible (application-specific) menu lists.

The hierarchical design of graphical windows in IDL enables each viewer to keep control over his state, hence to control *data flow*, and to be realized several times with different data selections.

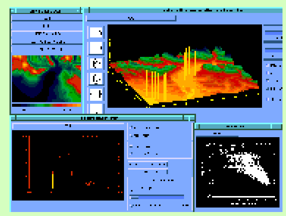
Two or more viewers may be *linked dynamically* and exchange spatial indices of data sub-selections; This design allows a direct comparison of data values in different viewers and multiple representations.

***DataScaping was implemented with ©IDL (Interactive Data Language) by RSI.

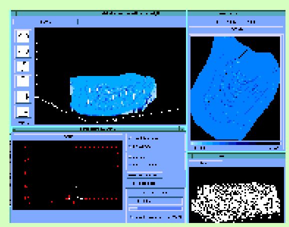
1rst example: yearly mean temperature data set (Switzerland, interpolated on a 1km grid). Upper left: 2d-image-viewer with interactive query of geo-reference and pixel values. Neighbor-values (lower left) and a user-defined profile (center right) are directly querried by mouse and results displayed in dynamically linked windows. The meta-information-viewer (upper right) showes the color mapping, statistical information, resolution and no-data-values of the temperature data set.



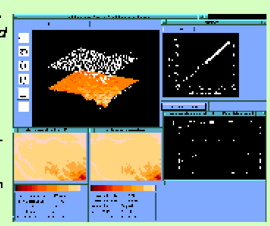
2nd example: exploring relations between a precipitation,temperature, and elevation data set (1 km grid) in multiple linked windows. The geographical locations of the selected values from the histogram plot are highlighted in the color-coded surface plot.



3rd example: DEM of the Murtel rock glacier (CH, 100m res.) and interpolated snow height. upper left: the interactive 3d-DEM was colorcoded by the snow-map (upper right) and original sampling points (white sticks) overploted. lower right: the DEM as wireframe surface may be linked with the histogram (lower left). Monthly snow maps can also be animated (ns).



4rth example: investigating two differently interpolated temperature data sets by comparing them in various graphical presentations. lower images: both data sets show only slight differences. upper left: the two-surface-viewer provides interactive slicing and displays dynamically both profiles in a seperate window which allows for more detailed comparison. upper right: scatterplot of both temperature data sets.



Environmental modeling involves large volumes of multivariate data. DataScaping is an experimental environment promoting interactive data visualization for multivariate data analysis, validation of modeling results and exploring relationships between spatially configured objects. Its concept is based on multiple representations of data, seamless integration of realistic and analytic graphical representations and dynamic linking of windows.

January, 1996