Title
Land Use Modeling Workshop

Authors
National Center for Geographic Information and Analysis
USGS

Publication Date
1997-06-05

The Land Use Modeling Workshop

June 5th to 6th, 1997

EROS Data Center

This Workshop was sponsored by:

USGS science for a changing world

NCGIA National Center for Geographic Information and Analysis

Introduction

The National Center for Geographic Information and Analysis (NCGIA) and the USGS sponsored a workshop on Land Use Modeling at the EROS Data Center in Sioux Falls, South Dakota from June 5th-6th, 1997.

Included in this document are:

- The meeting summary
- A list of participants and the steering committee
- Participant statements
- Full position papers from participants, where applicable

Meeting Summary

THE STATE OF THE ART OF LAND USE MODELING

I. Current State of the Art in Land Transformation Modeling

- Models are either stochastic (logit, markov, cellular automata, etc.) or processed based (dynamic ecosystem model).
- Models are developed independently by various teams with essentially no reuse or interoperability components.
- Models are linked to a heterogeneous group of external applications, such as GIS, statistical packages, and visualization packages.
- A few models exist which utilize distributed computing, but are application or model specific.

II. Projected State of the Art in Land Transformation Modeling

- Models are linked stochastic, process based, and other (fuzzy expert system, individual based, etc.).
- Integrated Problem Solving Environment (PSE) will be widely used.
- The PSE will incorporate a uniform interface to external applications.
- Transparent distributed computing will be widely available.
- Data interoperability is widely supported.
- Uniform GUI - Learning environment and modeling interface.

COMPARATIVE MODELING

Land use transition models, while reflecting a varied heritage and disciplinary background, share many commonalities. Common approaches are the use of transition probabilities in a class transition matrix, the use of multinominal logit methods, cellular modeling, and the use of the GIS weighted overlay approach. Emerging models that show potential are object oriented, expert system assisted, and use cells as their spatial unit rather than parcels.

Model choice issues related to source data include the choice of data model (features, cells or objects), and the comparability of multitemporal land use maps. Also important is the spatial aggregate structure of the modeling unit, be it watershed, county, metropolitan area, etc. Of almost unanimous concern in land transition modeling are the stability, estimation of, and autocorrelation in the transition probabilities.

Research on comparisons between models, such as that conducted on global climate models, is important to evaluate model performance, and to assist in validation.

DATA ISSUES AND NEEDS

Land Use and Land cover information is essential for the calibration, baseline, and testing of land use models. The information needs to be continuously updated, at spatial and temporal resolutions that are compatible with the models and have appropriate thematic content and detail.

Issues:

Are discrete categories appropriate? Spatial boundaries are occasionally more appropriately represented as fuzzy gradations. However, land use boundaries are typically fairly crisp because they are based on ownership. However, probabilistic models might benefit by maintaining probabilities as more continuous representations of land cover/use. Further, because aerial and satellite image processing provides much of the data, probabilistic or fuzzy approaches might be used. More continuous data might also be used and represent land use intensity, and changes in intensity.

Time and space scales are co-dependent and critical. Higher spatial resolution data are typically constrained to less frequent updates. For example, land use/cover information from Landsat TM might be updated over a large area every 5 years or decade (e.g. MRLC) whereas data at a resolution of AVHRR might be updatable monthly. Generally, detail as well as quality decreases with age of data.

Remote sensing provides raster data. Is that appropriate or sufficient? What land use modeling approaches use grids vs. networks vs. objects as data models? These issues may be spatial and process-scale dependent. For example, very detailed models often work with individual entities (plants, people), whereas more aggregate models work better with a grid framework.

Use vs. Cover. Some of the models predict land use change based on economics, some predict land cover based on biophysical processes. Yet, the difference between use and cover are often not distinct for the purposes of data collection. Anderson's classification attempts to combine use and cover. Satellite remotely sensed data provide land cover information well, but land use information requires more spatial detail and interpretations.

Needs:

Rethink Anderson - Do we need a new classification system? In what way? Can Anderson be improved? Do land use and land cover need separate systems? Can they be amalgamated?

Landsat is critical - provide support for program.

There are more data choices coming online with new private and public sector satellites. Modelers should be aware and flexible to make use of the most appropriate resolution data.

Monitoring and Rapid Assessment Protocols. We need to develop methods to quickly monitor and observe changes on large scales.

Is MRLC model appropriate?

APPLICATIONS

Applications are problem specific, tied to a geographic hierarchy. The spectrum of analysis ranges from global to local issues and areas of interests. Land use analysis models need to operate at the different geographic levels. The resolution of the data should be transparent to the model.

A land use model is applied to generate the results of a proposed scenario. A scenario usually has a focus which has either an urban or ecological driver(s). A focus or driver is comprised of factors, parameters or criteria that the user of the model can alter, prioritize, or weight to serve alternative scenarios. The land use model should allow the user to vary the time frame for each scenario being generated. The results of the scenario or application can be used as output to launch another model, to calibrate the land use model, and to assist in the decision making process.

POLICY/SOCIAL/ECONOMIC FACTORS

This is a very complex subject area so no single consensus can be reached in a short session. A list of topics/concerns emerged. This is a summary of some of the elements under ten of these.

1. Policy, social and economic (human dimension factors) are indispensible to models of land use change and, therefore, land cover.
2. The indicators that are used to record or measure the factors to be entered influence very much how many converging (confounded) factors are or are not included.
3. The united variables that are selected often shapes (or reshapes) the way the system is understood.
4. The assumptions about human behavior, human choice, and system stability and predictability determine the degree to which the model is simple and manageable but removed from reality or complex and unwieldy but approximating reality. The assumptions must be stated clearly and challenged regularly.
5. The incorporation of human dimension factors means necessarily that other disciplines are

involved. They bring both different assumptions (different indicators) and different methodologies. These complicate interaction but open doors for enhancing model performance (for example should multinominal logit be used where probit analysis can be used?)

6. In the system we wish to model, experimentation is not possible and much of the bases of reductionist science is challenged.
7. Scales of space and time: The processes of human activities operate at very different time scales from ecological processes with which they are linked. The local spatial scale of landscape units are affected by social, economic and policy factors at smaller scales (larger area).
8. Surprise in the system is inevitable and all systems are ultimately idiosyncratic. This raises the question of the value of models that purport to be general. At one scale, individual decisions are determined by very personal circumstances and quality of life considerations that are hard to measure. At another scale, things like earthquakes, fires, war, economics cycles, are unpredictable but leave indelible marks in land use processes.
9. Therefore, for all of these reasons it seems as though it might be futile to attempt to model. But such models are necessary because they (1) consolidate data, (2) make assumptions explicit, (3) permit testing of hypotheses, and (4) allow a basis for prediction about change.
10. But to the extent that the resultant information becomes a part of the decision making process, the model is a part of the process it is modeling and so the role of modeling as a means of shaping the future (rather than just predicting it) needs to be considered.

RESEARCH NEEDS

What are the big science, blue sky questions (requiring the expenditure of tens of millions of dollars) regarding land use and land cover modeling that can be answered through an organized research effort?

1. Are there a common (and limited) set of factors (variables) which explain (across a sample of the largest urban areas):
    - the extent and rate of urbanization/land cover change
    - composition of urbanization/land cover change.

   Can we identify them?
   Does data and modeling scale matter to the identification of these factors?
   How do we validate these factors (predictive validation, cross-validation)?
2. Which dimensions of scalability matter the most?
    - grid size

- polygon grain
- grid shape
- temporal scale (noise vs. signal)

3. Across a sample of representative urban regions, does the extent and composition of urban land use forms really affect the differential production of:
    - ozone and greenhouse gases (vs. meteorological conditions)
    - particulate (vs. Greenhouse gases)
    - watershed nitrogen and nutrient loadings (vs. hydrology) Are there critical periods that deviate from the typical?
4. What land use/land cover/landscape ecology primitives/objects would be most appropriate for a next-generation object-oriented land use/land cover modeling system. How do these differ from current conceptualizations of time, space, and scale in GIS?
5. What are the key public policy issues that land use/land cover modeling can usefully inform (i.e., shape appropriate interventions). At what scale? Can we develop some useful heuristics/screening criteria for answering this question?

FUTURE COLLABORATION AND INTERACTION

Multiple pathways should be used to collaborate and interact within and without the group represented here.

INTERNAL

Internal communication can be sustained most easily and consistently using a common web site. The site begun for this meeting could be continued for the purpose, perhaps using "forum" software so that all participants can post messages/information directly and immediately to the site. The contents of this user-maintained web site could then periodically be edited or distilled.

A follow-up to this meeting could be convened in a year or two. Timing could capitalize on reconvening of this group under current sponsorship, other joint funding opportunities, and the International GIS/Modeling Conference.

EXTERNAL

The above web page will be useful outside the group as well but any directly interactive function will probably be less important. Links to related pages will perhaps be more important.

A paper collection could be completed either as an NCGIA in-house report or a special issue of a journal. The first would not be peer-reviewed. Papers from this meeting could be used in either case with moderate adaptations.

A review paper could be done by joint authors as an article, a USGS open file report, or the front piece to a book which would give a systematic treatment, perhaps with case studies. A book could indicate the state of the art most comprehensively, particularly at this "moment of reinvigoration" and expansion.

SPECIAL

A "working group on land use modeling" could be formed as a virtual entity with the advantage that the burden of participation could be limited by individual desires yet its existence can be cited to endorse or sanction collective actions.

NOTES

Gray literature and intensive web page development are methods which do not have strong incentives attaching.

A writing project was begun at Santa Fe by Pijanowski, Parks, and Maxwell. A potential framework has been discussed and participation is invited.

A fourth GIS/Modeling Conference (following Santa Fe) is under consideration using a modified approach. Time, place and form are undecided.

Meeting Program

Training Facility, Introduction and Welcome
- Don Lauer, EDC
- Keith Clarke, NCGIA

Agenda and Goal for the Meeting
- Len Gaydos, EDC

Presentations: Land Use Modeling Research

Summary of Presentations: Themes and Issues
- Clarke & Gaydos

Discussion

Data Needs and Interests of Agencies, Funding Opportunities

A National Land Use Modeling Research Agenda: Progress and Challenges
- Ten minute presentations by panel.

Writing group "State-of-the Art in Land Use Modeling"

Steering Committee

Keith Clarke
NCGIA

Don Lauer
USGS EROS Data Center

Len Gaydos
USGS EROS Data Center

Bruce Quirk
USGS EROS Data Center

Participants

Michael Berry
University of Tennessee

Daniel Brown
Michigan State University
Department of Geography

Burstein, Steve
Middle Rio Grande Council of Governments

Keith Clarke
University of California
Department of Geography/NCGIA

Michael Crane
U.S. Geological Survey
Denver Federal Center

Thomas DiNardo
U.S. Geological Survey

Timothy Foresman
University of Maryland Baltimore County
Department of Geography

Len Gaydos
U.S. Geological Survey
NASA Ames Research Center

Jacqueline Geoghegan
Clark University
Department of Economics

J. Morgan Grove
USDA Forest Service
Northeastern Forest Experiment Station

David Hester
U.S. Geological Survey
Denver Federal Center

Laura Jackson
U.S. EPA

Gary Krizanich
US Geological Survey

John Landis
University of California
Department of City and Regional Planning

Donald Lauer
EROS Data Center
U.S. Geological Survey

John Lewis
McGill University
Department of Geography

Thomas Maxwell
University of Maryland

Thom Meredith
McGill University
Department of Geography

Carol Mladinich
U.S. Geological Survey
Denver Federal Center

Brad Parks
University of Colorado
Cooperative Institute for Research in Environmental Studies

Bryan Pijanowski
Michigan State University
College of Natural Science

Christopher Potter
NASA-Ames Research Center
Ecosystem Science and Technology Branch

Bruce Quirk
EROS Data Center
U.S. Geological Survey

Denice Shaw
U.S. EPA

Ashton Shortridge

University of California
Department of Geography/NCGIA

Jeffrey Spooner
US Geological Survey

Michael Stier
U.S. Geological Survey
Denver Federal Center

Daniel Sui
Texas A&M University
Department of Geography

Hanquin Tian
The Ecosystems Center
Marine Biological Laboratory

Janet Tilley
U.S. Geological Survey

Eric Wood
EROS Data Center
U.S. Geological Survey

Ming Zhang
University of California
Department of City and Regional Planning

Statements of Interest

Distributed Land-Cover Change Simulation and Multidimensional Interpolation

M.W. Berry
K.S. Minser
Department of Computer Science
University of Tennessee

Statement of Interest:

Computer simulations are used in landscape ecology to simulate the effects of human land-use decisions on the environment. Such decisions are influenced by both ecological and socioeconomic factors which can be represented by spatially explicit multidisciplinary data. With support from the U.S. Man and the Biosphere (MAB) program, we have developed (through a collaboration with several ecologists, economists, sociologists, and foresty personnel) the **L**and-**U**se **C**hange **A**nalysis **S**ystem (or **LUCAS**) for the study of land-use effects on landscape structure in such areas as the Little Tennessee River Basin in western North Carolina and the Olympic Peninsula of Washington state. These effects include land-cover change and species habitat suitability. Using a geographic information system (GIS) to store, display and analyze map layers derived from remotely sensed images, census and ownership maps, topological maps, and output from econometric models, a parallel/distributed version of LUCAS (pLUCAS) was developed for simulations on a network of workstations. Targeting distributed computational environments reflects the resources available to most land-use planners, forestry personnel, and wildlife managers. We have recently conducted a formal performance evaluation of two pLUCAS distributed models on an ATM-based network of 12 SUN Ultra-2 workstations. Speed improvement factors as high as 8 (relative to serial runs on a single SUN Ultra-2 workstation) have been obtained using the PVM or MPI message-passing environments.

As part of the Integrated Modeling Project (IMP) of Southern Global Change Program, we are responsible for the development of the (second) IMP module which facilitates the horizontal integration of forest responses to environmental stresses and disturbances through the use of micro-scale cellular automata. This module is being developed from the LUCAS modeling system prototype. Stochastic attributes used by LUCAS will incorporate the frequency distributions of output results generated by the IMP's Linked Dynamic Model. Overall focus of the Integrated Modeling Project (IMP) is to integrate forest health and productivity assessments of southern and southeastern forests by taking into account changing climate, air quality, and land use changes.

In order for LUCAS to aggregate site index and forest growth types (height vs age) attributes to forest stands across the southern region, response surfaces must be developed from selected outputs of the codes comprising the Linked Dynamic Model. Using 5-dimensional interpolation based on the

Modified Shepard's Method developed by Robert Renka (Univ. of North Texas) we are building a portable object-oriented (C++) software package that will produce interpolated values for the

1.mean,
2.standard deviation,
3.minimum value,
4.maximum value, and
5.distribution type (normal, lognormal, uniform, etc.)

of any model output (e.g, site index or the height of of dominant trees at age 25) based on the following independent variables (or conditions):

1.atmospheric carbon dioxide,
2.ozone exposure,
3.nitrogen deposition,
4.temperature, and
5.precipitation.

The *hypervolumes* or response surface data produced from the proposed activity will be constructed using the netCDF (network Common Data Form) which is a popular machine-independent format for representing GIS and other scientific data. These hypervolumes will be used to integrate the influence of environmental factors on the forest conditions modeled by the codes in the IMP's Linked Dynamic Module. These hypervolumes may also be of great use by other researchers (outside of the IMP project) in the study of forest growth and production in the southeastern US.

In addition to the development of the 5-dimensional hypervolumes for the IMP driving variables, we will be porting the current MPI-based pLUCAS prototype to a newly acquired IMP SP/2 (total of 40 processors). Cycles on this machine will be available for future pLUCAS-based simulations associated with Module II of the IMP project.

Credits:

Our research in **computational ecology** has been supported by the Southeastern Appalachian Man and the Biosphere (SAMAB) Program under U.S. State Department Grant No. 1753-000574 and University of Washington Subcontract No. 392654, by the National Science Foundation under grants NSF-ASC-94-11394 and NSF-CDA-95-29459, and the USDA Forest Service under Contract Nos. 29-1286-96 and SRS-CA-96-067.

References:

M. W. Berry, R. O. Flamm, B. C. Hazen, and R. L. MacIntyre. Lucas: A System for Modeling Land-Use Change. **IEEE Computational Science and Engineering**, 3(1):24-35, June 1996.

B. C. Hazen. A Distributed Implementation of the Land-Use Change Analysis System (LUCAS) Using PVM. Master's thesis, University of Tennessee, Knoxville, August 1995.

B. C. Hazen and M. W. Berry. The Simulation of Land-Cover Change Using a Distributed Computing Environment. **Simulation Practice and Theory**, 1997. In Press.

R. J. Renka. Multivariate Interpolation of Large Sets of Scattered Data. **ACM Trans. on Math. Soft.** 14(2):139-148, June 1988.

A Study of Decadal Scale Land Use/Cover Changes Across the Upper Midwest: Implications for Land Use Change Modeling

Daniel G. Brown
Assistant Professor
Department of Geography
Michigan State University

Current Research

A research group at MSU (of which I am PI) has recently been awarded a three-year grant under NASA's Land Cover and Land Use Change (LCLUC) program to study and model the relationships between socioeconomic and demographic changes and changes in the types and patterns (i.e., fragmentation) of land use and land cover across the Upper Midwest between 1970 and 1990. This project builds on two on-going projects. One is a collobration with the USDA Forest Service that is examining increased fragmentation of land ownership throughout the same region. The other, headed by Dr. Bryan Pijanowski (another workshop attendee) among others, involves the devlopment of a GIS-based spatial model of land use change. My interests are not in describing the model, but rather some of the issues involved in the interplay between the model and a data-rich empirical investigation on a regional scale.

The two previous projects:

1. describe changes in patterns of land ownership at 136 sample sites (selected through stratification by county type and location within 17 sample counties; Figure 1);
2. examine relationships between land ownership fragmentation and socioeconomic factors; and
3. model land use changes using a GIS-based, object-oriented process model.

The objectives of the LCLUC project are as follows:

1. collect, scan, and interpret land use from aerial photography from the 1970s, 1980s, and 1990s for 136 sample sites (three by three survey sections in size) throughout the Upper Midwest.
2. collect, classify, and mosaic land cover from North American Landscape Characterization (NALC) MSS triplets, 1970s, 1980s, and 1990s, for the entire region.
3. quantify the relationships between socioeconomic drivers and land ownership fragmentation, land use changes and fragmentation, and land cover changes and fragmentation.
4. model rates, types, and patterns of land use and cover changes at multiple spatial scales using a modified version of the Land Transformation Model (LTM) of Pijanowski and others (In Review).

Figure 1. Location of study region (not shaded), sample counties (in blue), sample sites (in red), and MSS scenes (outlined).

I am also working on a project recently funded by the National Science Foundation to compare observed spatial patterns of vegetation at treeline with the spatial patterns that result from a plant growth model (in the JABOWA-FORET tradition, but with some modifications to incorporate spatial processes). In this study we will attempt to predict the spatial pattern of productivity, using the leaf area index (LAI) as one measure. Both of these projects are just now getting started. What follows are thoughts that have arisen from my involvement in the planning stages of these two projects.

The Relationships Between Land Use Change Models and Data

"Nobody believes a model save its developer, everybody believes a data set except its collector"

-- Anonymous

I recently read this quote in the proposed framework for the National Environmental Monitoring Initiative by the National Science and Technology Council (NSTC) Committee on Environment and Natural Resources. It suggests something about the relative roles of data and models in science. In reality modeling and observation are two completely different endeavors. Yet they are dependent on one another in completing the cycle of scientific inquiry.

Reconciling models and data can be difficult. One illustration in the context of land use modeling can help make the point. Multi-spectral satellite sensors are typically proposed as fairly consistent sources of data with which land use change models can be compared. Most remote sensing specialists know that obtaining land *use* information from satellite data with a resolution the Landsat sensors (TM or MSS) is nigh to impossible (see Figure 2 below). Satellite data are good at providing information on land *cover*, not land use. However, modeling efforts tend to focus on land use because it is more directly a product of the social value placed on the land. Land cover is a consequence of how the land is used.

Figure 2. Color Infrared photo from Crawford County, Michigan (resolution = 2 m). Residential areas bounded by the road network in the subdivision on the left is spectrally very similar to the forest on the right. A spectral classification would most likely identify some of this area of residential USE, with the exception of the roads, as forest COVER, especially when aggregated to 30 or 60 meters.

Although difficult, reconciliation of data and models is necessary. Observational data are necessary if one is to test the performance of a model or to calibrate any unknown parameters. Also, a t0 map is typically needed to set a base line for the model. The source of the baseline map is typically remote sensing. Furthermore, the observation of changes without the structure of a model can be very difficult to generalize.

How can mapped land data be reconciled with land use models and vice versa? I can think of four different modes for such reconciliation, most of which have already been used to some extent. Although these are not new approaches, I think it useful to structure our thinking about how data supports modeling and vice versa. *First*, one can simplify the problem of change to one that involves only land use types that are expressed on the landscape as easily distinguishable land cover types. This is the approach taken in models that address urbanization (i.e., the conversion of land from rural to urban uses) as the only change (Clarke et al., 1996; Pijanowski et al., In Review). Unfortunately, all change is not urbanization, nor does urbanization always look the same (see Figure 2).*Secondly*, models can be compared and constructed with maps of land use (Geoghegan et al., 1997). Land use mapping is reasonable over a small area, but on a regional or global scale can be an overwhelming task because (a) data at the resolution of Landsat or coarser are insufficient for mapping most land uses and (b) automated multi-spectral algorithms are ineffective for identifying land use. *Third*, an attempt can be made to model land cover changes (Turner, 1990). This is usually done through the use of empirically based approaches (i.e., Markov transitions) which tend to be limited by the low level of explanation in the model. Therefore, these models have more serious limitations on how far they can be extrapolated in space or time. *Finally*, a model might attempt to explicate the linkage between land use, which is a product of many social driving variables, and land cover, which is a product of those driving variables and the resulting land use.

I am interested in exploring the degree to which different land uses result in certain types, proportions, and patterns of land cover. This requires information on both land use and land cover. In our study, we will map land use from aerial photography and parcel boundaries within our 136 sample sites, and land cover region-wide from MSS data. A modeling question becomes how to model land cover change regionwide as a consequence of land use change, for which only a sample of sites is available. Linkages might also be developed to predict changes in other ecosystem properties (e.g., green leaf biomass, net primary production, and leaf area index) as a consequence of land use change. These properties can provide information about the biogeochemical cycle implications of land use change and can also be monitored.

References

Clarke, K.C., Hoppen, S., and 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.

Geoghegan, J., Wainger, L., and Bockstael, N., 1997, Spatial landscape indices in a hedonic framework: an ecological economic analysis using GIS.*Ecological Economics*, In Press.

Pijanowski, B.C., Long, D.T., Gage, S.H., and Machemer, P.L., In Review, A spatially explicit land transformation model: integrating policy, socioeconomic, and environmental factors. *Ecological Modelling*.

Turner, M. G. 1990. Landscape changes in nine rural counties in Georgia. *Photogrammetric Engineering and Remote Sensing*, 56(4): 379-386.

Statement of Interest

Stephen Burstein
Middle Rio Grande Council of Governments

The MRGCOG is assembling an inventory of existing land uses throughout the Regional Plan Study Area. An 18 category schema of land uses was developed for application to polygons generally larger than 1 acre. We are using a variety of sources and techniques to develop this coverage that is so essential to the LAM.

The MRGCOG and the USGS National Mapping Division are coordinating activities in their respective projects of the Regional Plan and the Middle Rio Grande Basin Project. David Hester of the USGS National Mapping Division attended most of a series of charette meeting to develop LAM. We are developing a Memo of Understanding to guide various data sharing and other joint planning tasks.

My interests in the Land Use Model Workshop are as follows: to learn more about a variety of approaches to modeling future land use in conjunction with other functional areas which may be applied to the model development on which we are working, to contribute to the "state of the art" of land use modeling, to work on a better joint vocabulary among geographers, urban planners, modelers, and other related disciplines on this topic, and to advance productive working relationships between regional and local planning agencies and the USGS Council of Governments (MRGCOG). He holds a Master of Urban and Regional Planning degree from the University of Colorado at Denver and a B.A. from the University of Michigan. He is a member of the American Institute of Certified Planners.

The purpose of the Regional Plan is to create a long range vision and strategy for managing growth and development within the Middle Rio Grande region. The vision and growth management strategy is intended to meet the goal of attaining sustainable, long range economic vitality in the region. The achievement of the goal will require the development of a long range (horizon year 2050) land use-based framework which encompasses the linkages between regional land use and transportation systems, and other interrelated regional functional planning areas (e.g., water, air quality, linked regional open space, utilities, economic development and housing).

The Regional Plan project is scheduled to require approximately four years. The project consists of eight overlapping tracks of activities: the start-up phase "tool box," public involvement, advisory committees, supporting studies, developing scenarios, testing of scenarios, selection of preferred alternative, policies and implementation plan. We are one year into the project. This planning activity will result in a region-wide development plan to be adopted by the MRGCOG Board of Directors for the use of governmental agencies, non-profit agencies and businesses committed to involvement in regional issues, and private citizens. Local governments will be able to use the plan as a regional systems "template" guiding local comprehensive plan. The plan will guide MRGCOG planning activities in such areas as growth management, transportation planning, and regional water planning. The long range regional transportation plan developed by MRGCOG for the Metropolitan Planning

Organization (MPO designated by the U.S. DOT), to be updated through a Regional Major Investment Study, will be consistent with the preferred land use scenario.

Other regions have engaged in similar regional planning efforts from which we have learned important lessons and in some cases borrowed methodology. Those regions include the Puget Sound Regional Council, Portland Metro, Denver Regional Council of Governments, San Diego Association of Governments, and Palm Beach County.

The study area for the regional plan consists of the four County region in State Planning and Development District 3, plus a 360 square mile portion of southern Santa Fe County encompassing Edgewood. The total study area contains approximately 9,600 square miles. The region includes the mainly contiguous metropolitan area in and around the principal city of Albuquerque, several communities along the Rio Grande to the north and south of the Albuquerque metropolitan area, and a number of rural farming and mountainous communities located more than 40 miles from the metropolitan area. Additionally, Reservation lands of ten independent Indian Pueblos, a portion of the Canoncito Navajo Reservation, a portion of the Jicarilla Apache Tribe Reservation and a portion of the Navajo Nation "Checkerboard" area are located within the region.

MRGCOG is developing a land use analysis model (LAM) to assist in developing alternative land use scenarios. It is anticipated that a long list of 8-12 scenarios will first be developed, followed by a short list of 4 scenarios, and finally a preferred alternative incorporating various refinements. The consultant firm of Planning Technologies, Inc. in Albuquerque, headed by T. Michael Corlett, has been working on the LAM since November of 1996. The computer application is being designed to operate on ArcView. Analyses to be performed by LAM include: (a) an ability to generate socioeconomic and land use forecasts for the regional plan development scenarios; (b) sketch planning tools to allow planners to quickly design land use schemes and transportation or other public works systems responsive to those scenarios, (c) ability to "launch" other models such as the EMME/2 trip assignment model, Mobile-5 air quality model, a wastewater flow model, and possibly the USGS water flow model.

LAM will consist of the following modules: Scenario Development Module to draw and edit land uses interactively on screen, Economic Structure Module allowing depiction of relationships between forecast growth (in terms of SIC codes) and the demand for lands of different use, Disaggregation Heuristics Module to forecast development patterns through the GIS depiction of various types of rules or heuristics, Disaggregation Module to perform allocation, or disaggregation, of regional control totals of forecast socioeconomic variables to and use polygons to absorb development, and Editing Module to alter and revise land use maps.

The MRGCOG is assembling an inventory of existing land uses throughout the Regional Plan Study Area. An 18 category schema of land uses was developed for application to polygons generally larger than 1 acre. We are using a variety of sources and techniques to develop this coverage that is so essential to the LAM.

The MRGCOG and the USGS National Mapping Division are coordinating activities in their respective projects of the Regional Plan and the Middle Rio Grande Basin Project. David Hester of the USGS National Mapping Division attended most of a series of charette meeting to develop LAM. We are developing a Memo of Understanding to guide various data sharing and other joint planning tasks.

My interests in the Land Use Model Workshop are as follows: to learn more about a variety of approaches to modeling future land use in conjunction with other functional areas which may be applied to the model development on which we are working, to contribute to the "state of the art" of land use modeling, to work on a better joint vocabulary among geographers, urban planners, modelers, and other related disciplines on this topic, and to advance productive working relationships between regional and local planning agencies and the USGS.

Statement of Interests

Keith C. Clarke

Department of Geography/NCGIA

University of California, Santa Barbara

Land Transition Modeling

My interest in land use change comes from a background in land use mapping, enriched by a visit to the National Mapping Division of the USGS in 1992. Before concentrating in work on Analytical Cartography, GIS and simulation modeling, I also researched and taught urban geography. In 1992, Len Gaydos and I worked together with others to develop a small land transition modeling group within the USGS global change research program. This project became the Human-Induced Land Transformations project. Primarily centered on the transition to urban land use in the San Francisco Bay area, we realized that to understand urban growth, one needed to know the path of historical land changes. We built a historical land transitions data base for San Francisco, and used the data to build and then calibrate a cellular automaton model of urban growth.

The success of the HILT project led to several sequels. At the USGS and elsewhere, a second data set was constructed for the Washington/Baltimore metropolitan region. The data were also animated successfully, to generate some very revealing visualizations, under a project called temporal urban mapping. As student of mine in New York, Jacob Kramer, applied the Cellular model to a small area in the Stirling Forest area. From my perspective, as the modeling team leader, I extended and refined the model developed in the Bay Area, and began a new project called Gigalopolis, funded out of EDC. This project is extending the model to cover land cover transitions.

The land use change work is that I hope to cover in the paper for this workshop. The work is based on an extended cellular model which (1) takes the cue for land use change from the amount of urbanization in the previous time period (2) uses artificial agents called "Deltatrons" to both diffuse and enforce change and (3) includes additional requirements for calibration and validation.

Issues I am interested in at the meeting are:

- How do land cover transition probabilities change over time?
- What are the major determinants of specific land cover transitions?
- Do land use models scale?

Can innovations in visualization and representation be used for understanding the results of model outputs?

USGS MIDDLE RIO GRANDE BASIN STUDY - STATEMENT OF INTEREST

Thomas Dinardo
U.S. Geological Survey

The U.S. Geological Survey (USGS) Middle Rio Grande Basin (MRGB) Study is a 5-year effort by the USGS and other agencies to improve the understanding of the hydrology, geology, and land-surface characteristics of the Middle Rio Grande Basin. The Santa Fe Group aquifer is the main source of municipal water for the region. The New Mexico State Engineer Office manages the water resources in the basin and has declared the basin a "critical basin"; that is, a ground-water basin faced with rapid economic and population growth where there is less than adequate technical information as to the available water supply. The USGS will study the hydrology, geology, and land-surface characteristics of the basin to provide the scientific information needed for land use and water-resource planning. The study began in 1995 and shall be completed in 2000.

The Middle Rio Grande Basin is the area within the Rio Grande Valley extending from Cochiti Dam downstream to the community of San Acacia. The study area contains approximately 3,000 square miles and includes the Albuquerque/Santa Fe urban corridor. In the next few years, the City of Rio Rancho will surpass the City of Santa Fe as the second most populous municipality in the State of New Mexico. The Bureau of the Census has projected that the region's population could swell by 55 percent by the year 2025. The Middle Rio Grande Council of Governments projects an additional 700K people will be added to the region for a total population of 1.5 million by the year 2050.

In 1995, the USGS published the results of a 2-year study of the Santa Fe Group aquifer, which serves as the main source of municipal water supply for Albuquerque and other communities in the Middle Rio Grande Basin. This earlier study simulated water levels in the aquifer through the year 2020 using four different water-use scenarios. The study found that the quantity of water in the aquifer available for municipal supply was significantly less than previously estimated. Because approximately 600,000 people (40 percent of the population of New Mexico) live in the study area, water shortfalls could have serious consequences for the State. Recent projections indicate that the Santa Fe Group aquifer may be depleted by the year 2060.

Ground-water withdrawals in the region are also causing other environmental problems such as land subsidence and contamination of existing wells due to nitrate leaching. As a consequence of sole reliance on ground-water for domestic water supplies, the City of Albuquerque has acquired 48,000 acre-feet of surface water rights known as the San Juan-Chama project. However, delivery of San Juan- Chama water to the City of Albuquerque will not occur until the year 2003. Even though the Rio Grande water is all appropriated and the San Juan-Chama water is not yet available, the City of Albuquerque continues to entice major water users such as microprocessor manufacturer, Intel, to the region without contemplating the long-term sustainability of those development decisions. Water rights adjudication is already occuring in the Lower Rio Grande region of the State and eventually the problem may migrate to the Middle and Upper Rio Grande Valley.

As part of the USGS MRGB Study, the National Mapping Division (NMD) plans to characterize the land surface of the region. NMD shall integrate thematic data layers in order to derive a composite land development attractiveness value. Researchers will investigate alternative land use development scenarios for expanding water needs for the timeframes of 2020 and 2050 as part of the land surface analysis. Historically, man's impact on the landscape has caused land cover to evolve into land use. Specific land uses are more conducive to ground- water recharge and conducting a trend analysis will allow researchers to determine the long-term impacts on the land surface. Consequently, NMD shall compile temporal LULC snapshots (i.e., 1930's, 1950's, 1970's, and 1990's) that will be used for trend analysis as well as to forecast future land use patterns. Predicting future land use demand shall be modelled by associating specific land use categories with the ability to absorb regional socioeconomic forecasts for population and employment growth.

Human Behavior and Land Use Models: The Patuxent Watershed, Maryland and Southern Yucatan, Mexico.

Jacqueline Geoghegan
Assistant Professor
Department of Economics
Clark University

As an environmental economist, my research interests are in the study and modeling of human behavior with respect to their natural environment. With specific regard to land use, my research is in developing theory-based models, which are then estimated with revealed human choice preference data. My work in land use modeling is in two areas: the Patuxent Watershed in Maryland (with economists and ecologists at the University of Maryland, funded by EPA & NSF) and the southern Yucatan peninsular region of Mexico (with geographers at Clark University, funded by NASA).

Human-induced land-use/cover change is being modeled for the data-rich Patuxent Watershed of the Chesapeake Bay, thus providing the spatial configuration and dynamic evolution of a landscape by capturing ecological functions, human behavior, and their interaction. The effort links remote sensed data on land use and cover with a variety of spatially explicit socioeconomic and physical data as well as "communicating" ecological and economic models. Each model employs a landscape perspective that captures the spatial and temporal distributions of the services and functions of the natural system and human-related phenomena, such as surrounding land-use patterns and population distributions. Configuration and reconfiguration of the landscape follows from the intertwining of these phenomena, and the Patuxent work offers the potential for a richer model of land use and its change by accounting for spatial heterogeneity and linking land-use conversion to features of the landscape. The aim is to predict the probability that a given pixel, of a given description and in a given location, will remain in its current use or be converted to an alternative use. While the conversion process is affected by inertia and other disequilibrium considerations and constrained by zoning and other land use controls, the change in land-use probabilities are functions of the value of the parcel in alternative uses. Consequently the analysis must be able to explain what factors affect land values in alternative uses.

The land uses in the 7,000 km2 of seven counties of the Patuxent Watershed located in Maryland, range from Washington, D.C. suburbs to rural and agricultural southern Maryland. The conversion of agricultural and forested land (open use) to residential uses constitutes 78% of the total land-use change in the seven counties of the watershed during the past ten years. As a consequence, the economic modeling effort focuses on the prediction of "open" land use to residential use through a four part process: (i) analysis of residential value as function of a variety of spatially related economic and ecological variables that are hypothesized to affect residential land values, estimated on actual transactions of residential parcels; (ii) use of the estimated coefficients of land value estimated from actual transaction to predict values for "open" land were it to be converted to residential use; (iii) use of these predictions with other explanatory variables such as zoning, soil type, and costs of conversion to estimate the spatial distribution of the relative probability that any such land will be developed; and (iv) linking these relative probabilities with a macroeconomic model of the state of the local economy to predict annual housing starts to therefore predict how many of the pixels will change in a given year.

The model is a utility-theoretic econometric model of human behavior affecting land use decisions, not driven by GIS-determinism. Using spatial data leads to interesting complications such as spatial autocorrelation, temporal dynamics, and spatial structural change. Because of these, applying standard econometric techniques to either aggregate or disaggregate spatial data generates nonsperical disturbances, misspecification, and measurement error. Therefore new estimation techniques in spatial econometrics have been developed to take some of these issues into account in the Patuxent modeling effort (e.g. Bell and Bockstael, 1997). It is still an empirical issue as to whether the information gained by using spatial econometric techniques vastly improves the estimation. The field of spatial econometrics itself is still in its early stages. The initial spatial econometric modeling work in the Patuxent model demonstrates the potential improvements in explaining and predicting land values (Geoghegan and Bockstael, 1995). Further improvements and refinements in both the theoretical and applied econometric modeling techniques for use with the Patuxent model are presently under way.

Another theme of this research agenda is to use the remote sensed data more creatively in land use modeling. For example, in order to better capture the spatial externalities that often characterize land use, and therefore have a major influence on land value, indices based on the diversity and fragmentation of the surrounding landscape around each pixel have been included in the Patuxent land value model to further explain residential land values. The intuition on including these variables is that increasing diversity might adversely affect aesthetics, but may have convenience value signifying the proximity of important work, shopping, recreation and institutional destinations, so therefore it is an empirical question over which effect dominates. Fragmentation might be considered more obviously undesirable. Holding diversity constant, increasing fragmentation signals a hodge podge of land uses. A high fragmentation index is synonymous with a checkered landscape, and the potential for large negative locational externalities. Confusion over the sign of expected effects may be very much tied to the issue of scale, another issue that increasingly is important for the social sciences to consider as discussed above. Preliminary estimation demonstrates that these additional GIS-created variables, measured at different scales, can add explanatory power to the Patuxent model of housing values (Geoghegan, et al., 1997). The nature and pattern of the land uses surrounding a parcel have an influence on the price, implying that people care very much about the patterns of landscape around them, supporting the belief that severe externalities exist in land use and in land use patterns.

Another fruitful means for addressing the human-environment relationship, especially where spatially explicit data are sparse or of coarse resolution, is to link empirical models derived from the remotely sensed imagery with theory-based models of dynamics of land-use/cover change. An example is work underway in the southern Yucatan peninsular region (SYPR) that merges remote sensed data-based Markov modeling approaches with field-and statistical-based models of the various land managers that are producing the signals registered in the imagery.

SYPR covers the Mexican states of Quintana Roo and Campeche, northern Belize, and northern Petan, Guatemala, from the Caribbean to Gulf of Mexico. The dominant semi-deciduous tropical forests of the region came under assault after a major highway was built through the center of SYPR, connecting the east and west sides of the peninsula. This road became the pathway for various new land-users: first slash-and-burn farmers on communally designated lands, followed private ranchers, NGO-sponsored rice projects, and more recently, biosphere reserves. Most of these changes coincide with Landsat imagery.

This imagery can be used to classify land-covers by various time periods, cross checked by field observations, and applied to standard Markov chain analysis. This approach, based on the assumption that the immediate past is the past predictor of the near future because of "stationarity" of the processes involved, uses transition probabilities of past states (e.g., land uses) to estimate future ones. Markov approaches have been used successfully by ecologists to assess land cover, and seem appropriate to explore regions with "low" levels of "chronic" change, such as the ephemerally used forests in much of the SYPR. Their applicability for multiple land-use changes is less certain, especially where stochastic processes operate as in SYPR.

Markov approaches can be made spatially explicit on a pixel by pixel bases and the transitions of each pixel weighted by its cover and use characteristics as well as the varying processes operating within it. Exploring such elaborations for SYPR revealed that by accounting for slope, elevation, and distance to nearest different land cover predicted more than 95% of the spatial variance in forest cover (Ogneva-Himmelberger & Turner 1996). The results for other transitions, such as those of various kinds on agricultural land, were not so successful, in part because the stationarity principle applied less well to them. Since the initial cultivators entering SYPR were smallholder subsistence farmers and since the dynamics of agricultural change among these kinds of farmers tend to track well with population density, the transitions of the pixels were weighted by this factor. The overall results for the SYPR were disappointingas anticipatedbecause so much of the land cleared for agriculture in the region involved large-scale, NGO projects that were not responding the production logic of smallholders. Eliminating these projects from consideration and focusing on those areas in the SYPR where smallholder subsistence remains, the simple addition of population density improved the predictive capacity of the basic Markov approach (Ogneva-Himmelberger & Turner 1996).

Indepth understanding of the various land-use/cover changes in the SYPR, however, reveals that the processes leading to so much of the changes in the 1970s and 1980s are no longer operatingthe petro-boom period with significant investment in large-scale agricultural experimentsand that new processes related to changes in tenurial institutions are currently affecting smallholder production and livestock production. The variation in the different types of land managers can be incorporated into this kind of analysis, operating from approaches outlined in the Patuxent study above, essentially "socializing" each pixel in the analysis according to the kind of land managers associated it. The stochastic factors per se cannot be model, but they can be introduced into the basic approach through scenarios. The larger point, however, is that working from models of land managers and from the RSD, hybrid approaches emerge that prove fruitful for near-term projections and monitoring of land-use/cover change.

References

Bell, K. and N. Bockstael. 1997. "An Example of Spatial Economic Modeling: Land Use Conversion in Howard County, Maryland". Paper presented to the Association of Environmental and Resource Economists at the Allied Social Sciences Association meeting, New Orleans, LA, January 1997.

Geoghegan, J., and N. Bockstael. 1995 "Economic analysis of spatially disaggregated data: explaining land values in a regional landscape." Paper presented to the Association of Environmental and Resource Economists at the Allied Social Sciences Association meeting, Washington, DC, January 1995.

Geoghegan, J., L. Wainger, and N. Bockstael. 1997. "Spatial Landscape Indices in a Hedonic Framework: An Ecological Economics Analysis Using GIS", Ecological Economics 22(3).

Ogneva-Himmelberger, Y. and B. L. Turner, II. 1996. "Markov Modelling Experiments of Deforestation in Yucatan". Paper presented to the American Association of Geographers NESTVAL Conference, Clark University, Worcester, Massachusetts, November, 1996.

Statement of Interests

Leonard J. Gaydos

U.S. Geological Survey, EROS Data Center, Ames Research Center

Metropolitan Landscape Transformations

I have always been interested in land use change. My first geography project in college was mapping the growth of 3 medium sized cities from topographic maps. My first project with USGS was helping Jim Wray and others map land use and land use change for a selection of U.S. cities.

Several years ago as the National Mapping Division of USGS was defining its role in the U.S. Global Change Research Program, I advocated a goal of monitoring, analyzing, and predicting patterns of landscape change. I proposed a prototype study with several other scientists called Human Induced Land Transformations (HILT). Our goal was to map patterns of major land use changes at the regional scale for the last 200 years, and make predictions for the next 100 years. Our initial focus was on urbanization in the San Francisco - Sacramento region.

Our assumption was that urbanization, though the product of myriad personal decisions, could be characterized as an organic process that responds to simple parameters and factors such as availability of land, topography, transportation, and existing settlement pattern. We sought a model that would approximate urbanization in a region as it might be seen from a vantage point in space over several decades.

Those requirements led to collaboration with Keith Clarke who was currently serving on the headquarters staff of USGS National Mapping Division. Keith had previously worked on a wildfire model while at NASA Ames the previous two summers. That modeling, based on cellular automata, had a rich heritage, including applications to urban growth by Batty and Longley.

To parameterize the model, colleagues William Acevedo and Cindy Bell constructed a temporal urban mapping database for San Francisco - Sacramento based on interpretations of Landsat, topographic maps, and other data. Those data formed the basis for popular animations of past growth patterns (mpeg file) that were well received by the media in the San Francisco region.

A second database was constructed by William Acevedo, Janis Buchanan, Tim Foresman, Janet Tilley, Susan Clark, and others for the Washington-Baltimore region.

Land use change is a process of immense interest to scientists, policymakers, and the public. Based on the success of HILT, USGS is now exploring the potential of a Metropolitan Landscape Transformations initiative:

At the turn of the century, America has become a metropolitan nation. Productive farmlands, wetlands, forests, and deserts that formed the America of 1900 have been transformed during the past 100 years into housing and employment for approximately 275 million Americans by 2000. The objective of Metropolitan Landscape Transformations is to trace these alterations to the American

landscape over the past 100 years, compiling and reviewing the record of land use change and its many impacts, and to anticipate the changes to come in the next 100 years. Models of land use change will be used to predict future metropolitan growth and its impacts, based on current and anticipated trends. This will give policy makers, scientists and the public the "future maps" upon which to begin a dialogue on the kind of American landscape we want in the years to come.

USGS MIDDLE RIO GRANDE BASIN STUDY - STATEMENT OF INTEREST

Dave Hester
USGS - Rocky Mountain Mapping Center

The U.S. Geological Survey (USGS) Middle Rio Grande Basin (MRGB) Study is a 5-year effort by the USGS and other agencies to improve the understanding of the hydrology, geology, and land-surface characteristics of the Middle Rio Grande Basin. The Santa Fe Group aquifer is the main source of municipal water for the region. The New Mexico State Engineer Office manages the water resources in the basin and has declared the basin a "critical basin"; that is, a ground-water basin faced with rapid economic and population growth where there is less than adequate technical information as to the available water supply. The USGS will study the hydrology, geology, and land-surface characteristics of the basin to provide the scientific information needed for land use and water-resource planning. The study began in 1995 and shall be completed in 2000.

The Middle Rio Grande Basin is the area within the Rio Grande Valley extending from Cochiti Dam downstream to the community of San Acacia. The study area contains approximately 3,000 square miles and includes the Albuquerque/Santa Fe urban corridor. In the next few years, the City of Rio Rancho will surpass the City of Santa Fe as the second most populous municipality in the State of New Mexico. The Bureau of the Census has projected that the region's population could swell by 55 percent by the year 2025. The Middle Rio Grande Council of Governments projects an additional 700K people will be added to the region for a total population of 1.5 million by the year 2050.

In 1995, the USGS published the results of a 2-year study of the Santa Fe Group aquifer, which serves as the main source of municipal water supply for Albuquerque and other communities in the Middle Rio Grande Basin. This earlier study simulated water levels in the aquifer through the year 2020 using four different water-use scenarios. The study found that the quantity of water in the aquifer available for municipal supply was significantly less than previously estimated. Because approximately 600,000 people (40 percent of the population of New Mexico) live in the study area, water shortfalls could have serious consequences for the State. Recent projections indicate that the Santa Fe Group aquifer may be depleted by the year 2060.

Ground-water withdrawals in the region are also causing other environmental problems such as land subsidence and contamination of existing wells due to nitrate leaching. As a consequence of sole reliance on ground-water for domestic water supplies, the City of Albuquerque has acquired 48,000 acre-feet of surface water rights known as the San Juan-Chama project. However, delivery of San Juan- Chama water to the City of Albuquerque will not occur until the year 2003. Even though the Rio Grande water is all appropriated and the San Juan-Chama water is not yet available, the City of Albuquerque continues to entice major water users such as microprocessor manufacturer, Intel, to the region without contemplating the long-term sustainability of those development decisions. Water rights adjudication is already occuring in the Lower Rio Grande region of the State and eventually the problem may migrate to the Middle and Upper Rio Grande Valley.

As part of the USGS MRGB Study, the National Mapping Division (NMD) plans to characterize the land surface of the region. NMD shall integrate thematic data layers in order to derive a composite land development attractiveness value. Researchers will investigate alternative land use development scenarios for expanding water needs for the timeframes of 2020 and 2050 as part of the land surface analysis. Historically, man's impact on the landscape has caused land cover to evolve into land use. Specific land uses are more conducive to ground- water recharge and conducting a trend analysis will allow researchers to determine the long-term impacts on the land surface. Consequently, NMD shall compile temporal LULC snapshots (i.e., 1930's, 1950's, 1970's, and 1990's) that will be used for trend analysis as well as to forecast future land use patterns. Predicting future land use demand shall be modelled by associating specific land use categories with the ability to absorb regional socioeconomic forecasts for population and employment growth.

Statement of Interest

Laura Jackson
U.S. EPA

The U.S. Environmental Protection Agency's (EPA's) Office of Research and Development (ORD) is seeking partnerships in land-use change modeling in order to project ecological vulnerability at fine spatial scales. We are focusing exploratory research and testing existing techniques in the mid-Atlantic region because of the availability of spatially and temporally extensive ecological monitoring data, and because of numerous interagency environmental assessment projects ongoing the same region. Our initiative addresses the question, "Where will projected land-use change most threaten ecological resources in the mid-Atlantic region?"

Our priorities for vulnerability assessment include ecological resources directly displaced by land-use conversion, and those indirectly impacted by increased quantity and toxicity of runoff and air pollution. Because of our need to associate land-use change with likely ecological effects, we require spatially-explicit projections with resolutions in the neighborhood of 100m or higher. Temporally, we would like to project to the years 2010, 2030, and 2050, with known levels of uncertainty.

The mid-Atlantic study region is comprised of all or parts of ten eastern U.S. states, and includes the entire Chesapeake Bay watershed as well as those for the Delaware River and the Albemare-Pamlico Sound. Land-use conversion is one, if not the, primary stressor for ecological systems across this vast area. However, our reviews to date of land-use modeling techniques have led us to believe that meaningful projections require an intimate familiarity with the features and patterns of small places--several counties at best. Therefore, we are planning to approach this initiative in two stages--a coarse regional overview, and an intensive study of selected subregions that we then infer to be most ecologically vulnerable.

Stage One
The coarse regional overview would utilize county-level data on natural resource markets, land values, demographic patterns, and other socioeconomic variables to project land-use in 2010, by county, for forest, pasture, row crops, and developed land. Alternative projections would derive from models under development by USDA and university resource economists. By overlaying ecological resources of greatest frailty or concern (wetlands, large blocks of contiguous forest, steep slopes, etc.), we would identify subregions that appear to be at the greatest risk ecologically for the year 2010. Selecting subregions for more intensive study will also require evaluating which areas would be best served by ecological risk assessment. Areas amenable to the prevention or mitigation of ecological degradation will be favored over those where remedial efforts seem less plausible.

Stage Two
In selected subregions, we would explore additional county, sub- county, and spatially-explicit status

and trends for issues including local environmental resources, employment centers, zoning, and infrastructure, to understand the forces that affect township and individual landowner decisions. Subregional projections would be based on the estimated economic decisions of private landowners as well as on the likely continuation of trends in infrastructure expansion. Efforts will also be made to incorporate major driving forces into the models that are beyond subregional boundaries. These forces would largely include natural resource issues such as global markets, new production technology, and national policy. Subregional projections to 2010, 2030, and 2050 would include alternative scenarios derived from varying our assumptions about local and external forces. We plan to test several modeling techniques currently under development in regional universities for single watersheds and single- and several-county areas. We may find that certain models are most applicable to subregions dominated by particular economies such as agriculture or mining, or by high rates of residential conversion. In this event, we may apply multiple models in the region in order to project watersheds or habitats of greatest ecological vulnerability to land-use change. We are very much interested in the land-use change initiative of the USGS EROS Data Center and the National Center for Geographic Information Analysis because of the ongoing project in the Washington-Baltimore area, and the possibility in the future of modeling large regions at high resolution. We are eager to learn from the expertise of these organizations, and hope to become partners in the exploratory research.

Statement of Interest

Gary Krizanich USGS

I am presently working with a multidisciplinary team on the Southeast Michigan Drinking Water Initiative. The project is one of four funded nationwide by the USGS under its new Drinking Water Initiative Program. The aims of the Drinking Water Initiative Program are to (1) assess the quantity and quality of source waters used for drinking water supplies, (2) evaluate the vulnerability of source water to key contaminants of public concern, (3) evaluate the utility and effectiveness of existing source water protection strategies and (4) report results and information in ways that easily communicate with end users.

Recent studies of ground water resources in Michigan by the Michigan Department of Community Health (MDCH), Michigan Department of Environmental Quality (MDEQ), and the U.S. Geological Survey (USGS) show that there are several counties where concentrations of arsenic in ground water exceed the U.S. Environmental Protection Agency's (USEPA) established maximum contaminant level (MCL) for drinking water of 50 ug/l. According to MCDH, arsenic concentrations exceed the MCL in more than 30 percent of 317 supply wells sampled as part of a 5-year hydrogeologic study of Huron County. Ground water in at least 10 other southeastern Michigan counties also contain arsenic in concentrations that exceed or are close to the MCL. These counties have a combined population of nearly 2.7 million people (1990 census) including all or parts of the cities of Detroit, Flint, and Ann Arbor. The presence of arsenic at elevated concentrations is the most significant limitation on future development of ground water supplies in the most rapidly growing part of Michigan.

An Open Geographic Modeling System

Thomas Maxwell, Robert Costanza

University of Maryland,

Institute for Ecological Economics

Abstract

Developing the complex computer models that are necessary for effectively managing human affairs through the next century requires new infrastructure supporting high performance collaborative modeling.

In this paper we describe an ongoing program to develop 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 multiple 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 support of science, education, and policy making on many levels.

STATEMENT OF INTEREST

Carol Mladinich
U.S. Geological Survey
Denver Federal Center

The U.S. Geological Survey (USGS) Front Range Infrastructure Resources (FRIR) Project is a 5year effort by the USGS to improve the understanding of the location and characteristics of natural aggregate, water, and energy resources that are vital to sustaining the Colorado Piedmont and its infrastructure. The project will provide the public and decision makers with objective scientific information to determine the rate that infrastructure resources are being preempted by conflicting land uses.

The project encompasses a swath approximately 30 miles wide along the Front Range from Cheyenne, WY to Pueblo, CO (112 1:24,000-scale USGS quadrangles). A demonstration area has been selected to develop data sets and methodologies. This demonstration area extends from the south end of the Denver Metropolitan region to just north of Fort Collins/Greeley (45 1:24,000-scale USGS quadrangles). Models will be developed that relate resource occurrence to quality, quantity, and availability based on environmental and land use impacts. The determination of landscape processes and conditions using historical landscape patterns will be used to forecast which areas are likely to experience future development and to determine the effects of the development on local infrastructure resources. The land will be characterized by the various data sets and techniques will be developed for determining landscape change.

Statement of Interests

Daniel Sui
Department of Geography
Texas A&M University

My interests in urban land use modeling started with two land use/land cover mapping projects using TM imageries and aerial photographs in the late 1980s. Toward the end of these two projects, I came to the conclusion that the potential of remote sensing will not be fully realized if the thematic information such as LU/LC extracted from remote sensing imageries are not put into GIS. So my interest shifted from remote sensing to GIS, especially the integration of multiple data layers using conventional cartographic modeling for land suitability analysis. Because of the ad hoc nature of cartographic modeling using overlay and buffer analysis, I started exploring how to improve the analytical capabilities of GIS since 1990. I have concentrated primarily on the integration of GIS with spatial analysis and modeling with applications in the socio-economic arena. Specifically, I worked on the integration of GIS with a modified version of Lowry model to simulate the urban development patterns for the city of Hong Kong; the incorporation of fuzzy logic and neural computing in cartographic modeling to better handle the ambiguities and vagueness in spatial decision making. I am also keenly interested in exploring the characteristics of the emerging new urban forms and what are the processes responsible for these new urban forms. I firmly believe that efficient and effective urban land use modeling efforts must be grounded squarely in robust urban theories and vigorous urban theories must be capable of explaining the new urban reality during the information age. One of the key issues I am currently working on is how the new telematics revolution will change the urban structures, functions and what kind of new policy initiatives we need to deal with urban issues in the information city. I have also been working on a holistic approach toward urban modeling in which new urban theories on information cities and new modeling techniques in non-linear dynamics are seamlessly integrated to simulate urban development. I am also interested in alternative conceptualizations of space and time and new computational implementation strategies for the development of the next generation of GIS.

Statement Interests

Michael Stier
U.S. Geological Survey
Denver Federal Center

I have been recently assigned to perform research for the National Mapping Division of the USGS pertaining to land surface analysis. My plan is to produce a product or process that would characterize the land surface in a unique way using primarily USGS data sources. Because of the close relationship of the subject matter being presented at the Land Use Modeling Workshop, I am very interested in hearing what other researchers have accomplished or are pursuing in the study of land use as it relates to land characterization.

Statement of Interest

Bryan C. Pijanowski
Dean's Office, College of Natural Science and
Department of Entomology
Michigan State University

Modeling the Drivers to, and Impacts of, Land Use Change

My interests are 1) in modeling the drivers to land use change and 2) integrating land use change to ecological and economic models. I am the principal developer of the Land Transformation Model, which is a GIS-based land use change model. The model was first developed, under a cooperative agreement between the US Environmental Protection Agency and the Consortium for International Earth Science Information Network (CIESIN), with an application to Michigan's Saginaw Bay Watershed located in Michigan. The Saginaw Bay Watershed pilot model currently contains 13 driving variables (most contain subdrivers) that are policy, socioeconomic and environmentally based.

I received my PhD from Michigan State University in 1991 in the area of ecological modeling. I conducted filed research on birds and mammals and developed ecological models based on game theory. I have also worked in Integrated Pest Management where I have participated in a National pilot program that attempts to slow the spread of the gypsy moth using the latest integrated pest management techniques. Several MSU colleagues and I have also received funding from NASA to participate in a National Earth Systems Science Education (ESSE) program for undergraduate and graduate students studying the human dimensions of global change using NASA MTPE data resources, systems science and computer technologies such as GIS.

Most of my research in the next few years will concentrate on furthering the development of the Land Transformation Model. Dan Brown (MSU), Mike Vasievich (US Forest Service) and I received funding from the NASA Land Cover Land Use Change (LCLUC) program to analyze and model socioeconomic drivers of decadal scale land cover changes in the upper midwest. We will incorporate new scale-depended socioeconomic economics drivers into the LTM as well as some land cover dynamic measures. Sheridan Haack (USGS), David Long (MSU), David Hyndman (MSU) and I will attempt to link the current Land Transformation Model to hydrogeologic and geochemical models. An historical analysis of land use change and water chemistry and hydrology within Michigan's Grand Traverse Bay Watershed's streams along with rigorous sampling will be conducted. I am also collaborating with two other groups who will apply the principles of the Land Transformation Model to land use change in non-Michigan environs. One is along the north coast of Alaska where the LTM will be linked to cumulative impacts of human activities and the other is in Zimbabwe where the LTM will be integrated with forest succession and forest hydrology models.

I would be interested in participating in discussions on:

1. how drivers to land use change differ globally;

2. lessons learned on linking land use change models to ecological and economic "impact" models; and,
3. available tools that can visualize 3-D dynamic processes.

Statement of Interest

Christopher Potter

Research Scientist

Ecosystem Science and Technology Branch

NASA Ames Research Center

Modeling Land Use Change and Ecosystem Processes in the Amazon Basin

I am a terrestrial ecologist by training. My main research interests are global change, biogeochemistry, and land use. I've been at NASA's Ames Research Center for the past six years developing simulation models that couple these three areas of interest. The main tools of this trade are super computers, remote sensing, GIS, field research, and (hopefully) some creative thinking.

My main interest with respect to this Land Use Modeling workshop is on coupling our ecosystem modeling knowledge to land use modeling at the regional scale. Our current geographic focus area is the Amazon Basin. I am considering a series of model-based questions that can best be addressed through joint research among several disciplines:

What are the predominant land uses in different part of the Brazilian Amazon?

How is land use and management changing in the Brazilian Amazon?

How does land management influence the abundance and distribution of pastures and secondary forest growth, and the balance between clearing and re-growth?

How do changes in natural ecosystem processes influence these land-use changes and management practices?

Our regional geographic information system (GIS) for the Brazilian Amazon serves as the data source of land cover, climate drivers, satellite greenness images, and soil properties for input to the NASA-CASA (Carnegie-Ames-Stanford Approach) model at a 8-km grid resolution. Simulation results already reveal regional effects of forest conversion on plant production potential and soil carbon content, especially in seasonally dry areas. These results are being used to formulate a series of research hypotheses for testing in the next phase of regional modeling, which ideally will include linkage to land use simulations.

The background for our work is global change. Humans are causing environmental alterations of planetary significance. One of the clearest signals of human impact is the rapidly rising concentration of greenhouse gases like carbon dioxide (CO_2) in the atmosphere as a result of combustion of fossil fuels and changes in land use. Other trace gases such as methane (CH_4) and nitrous oxide (N_2O) can significantly influence the energy balance of the Earth. Moreover, these compounds are linked to atmosphere-biosphere feedbacks (CO_2 fertilization effects on vegetation), tropospheric chemistry (CH_4 reactions with OH, O_3 and NO_x dynamics) and stratospheric chemistry (N_2O-mediated destruction of ozone).

By the use of recently assembled satellite images of the global land surface, we have developed the simulation model called NASA-CASA to study of the role of terrestrial plants and soils in the cycling of carbon to and from the atmosphere. This research has produced a dynamic and detailed picture of the contemporary balance between photosynthetic fixation and microbial respiration of CO_2. In the NASA-CASA Biosphere model, a greenness index from satellite sensors is combined with modeled climate stress to estimate plant production. Carbon and nitrogen fluxes from decomposition of plant residues are simulated in the soil and at the surfaces of forest and grassland soils.

The results of our latest multi-year simulation of global ecosystems directly infer the presence of a net sink for carbon dioxide in the terrestrial biosphere over the period 1985-1988. This is, to our knowledge, an original research result that has not been demonstrated before at the global scale using actual remote sensing observations and land surface climate data. By combining medium-term (decade long) climate data with ecosystem modeling, new answers emerge with respect to biosphere-atmosphere exchange of trace gases resulting from climate fluctuations. However, the results from this ecosystem modeling study, along with others that have used satellite observations to infer plant production, must be qualified to point out that conclusions cannot be extrapolated directly to infer long-term (several decades) response of the biosphere to global warming or land use change. The processes of disturbance and land management must be included in additional modeling analyses that include potential geographic shifts occurring over several decades in plant functional types, physiological responses, and soil carbon turnover rates.

Statement of Interest

J. Morgan Grove, Research Forester
USDA Forest Service
Northeastern Forest Experiment Station

I have four areas of interest in land use modeling: 1) a social ecological approach to land use dynamics (i.e. land use change); 2) concepts of physical, biological and social differentiation of landscapes; 3) a hierarchical approach to land use dynamics, and 4) types and rates of land use dynamics. I hope these may be of interest to you as well.
1) A social ecological approach to land use dynamics

Researchers are focusing increasingly on an integrated framework for studying human ecological systems (figure 1). Many of the components of this framework are central to studying land use dynamics. How can such an integrated framework be used to enhance our understanding of land use dynamics?

2) Concepts of physical, biological and social differentiation of landscapes

How can an integrated framework be applied spatially? In some sense, land use is a measure of how landscapes become spatially heterogeneous. This spatial heterogeneity has a physical, biological and social component and these components interact with each other and change over time. What concepts of physical, biological and social differentiation can be used in order to understand land use dynamics?

3) A hierarchical approach to land use dynamics

Physical, biological and social differentiation occur at different spatial scales (figure 2), which suggests a multi-scale approach to land use dynamics. In this context, broad land use types such as urban/rural or forest, agriculture, residential, commercial / industrial may be associated already with different types and extents of land cover. But differences in human behavior are associated in predictable ways with differences in the social characteristics of people at smaller units of analysis (e.g. community, household, individual). For instance, in order to understand the relationship between the hydrology of a watershed at different land use scales, we may want to know how differences in human behavior (e.g. forest/vegetation management) at these lower scales affects the extent, distribution, structure, species diversity, and rates of regeneration, growth and mortality of forests and vegetation over time as well as inputs of fertilizers, pesticides and toxins into the watershed.

4) Types and rates of land use dynamics

While different types of land use change may occur at various spatial scales, there may also be different types and rates of land use change. For instance, there may be positive or negative feedback loops, time lags or non-linear rates (thresholds) of land use change. Also, some types of land use change may occur over a day or year, while other types may take centuries. How can we include explicitly an understanding of different types and rates of land use dynamics in our research?

Statement of Interest

Thomas Meredith,
Department of Geography,
McGill University

Contents

1. Research Interests

My academic interest is in the conservation and management of ecological resources with a particular focus on environmental impact assessment and community response to environmental change. Working with Parks Canada, I became aware of the need for new approaches to ecological conservation management. Through my PhD in plant population ecology and subsequent work in applied conservation, my research has been based on recognition of a need for solid science as well as solid linkages with the community of resource users. This arises from the belief that effective environmental protection will come not solely from either draconian legislation or land sequestration, but rather from an aggregation of local environmental systems that involve human communities in sustainable relationships with their biophysical resource base. I have worked with native communities in Quebec (the Naskapi and Inuit), as well as with rural resource communities in Quebec (La Mauricie), British Columbia, Mexico and Africa.

My link with land use modelling is that I am interested in the "cybernetics" of change within community-environmental systems and am attempting to determine how geographic information and spatial analysis procedure can best support community decision-making processes. This is work-in-progress. To me the particularly interesting element of modelling land cover change is a paradox discussed by the Canadian Global Change Program panel on "Critical Environmental Zone." That is: if models of land use change are both accurate and relevant, they will modify human behaviour so that the predictions based on the model are wrong. In other words, if models of change are simply based on historic probabilities of change, they fail to recognize the capacity of "the system" to learn and modify those probabilities. Models can be seen as having two purposes, one is simply to describe from the outside what is going on (and so the model is not really a part of the system) and the second is to provide information about futures so that undesirable futures can be avoided (and so, in this case the model IS a part of the adaptation process it is describing because it will modify actions). In conservation work, the second category of model is of most interest.

The question I would pose is how do models incorporate learning from the models?

In a Mexican project we assumed continued rates of land cover conversion and demonstrated what the landscape might look like in 10 and 20 years given trends of deforestation and urbanization. The intention was to modify land use decisions that will affect land cover change by demonstrating a model that, in effect, would prevent what it predicted. This exercise was based on an assumption that the community assigned a value the forest and would make appropriate decisions if the temporal scale of land cover change was made visible. But in this case, no attempt was made to quantify factors that would influence the probability that evidence of change in the environment would lead to change in human behaviour.

By contrast, in a project in Tanzania, a graduate student and I have been trying relate very specific use values for environmental resources associated with land cover types. For example, what is a given species of tree used for, how important is that use and what substitutes are there for it? That quantification of value would be a weighting that would influence the probability of compensatory action being taken by human actors in the landscape once predictions of land cover change were made known to them. The following is an abstract for a paper we are preparing on this work.

2. Abstract

Using value estimates for woody species and information about land cover change to identify conservation priorities in Northern Tanzania. (Draft).

Wynet Smith, T.C. Meredith, and T. Johns

Identifying conservation priorities in communities dependent on ecological resources requires consideration of local use patterns, the extractive impacts of local uses, and of the individual species availability and distribution. Merging of these information sets can help to identify locally valued species and species under pressure. Use-value is defined as a function of number of uses and of availability of substitutes. Conservation priorities can be expected to be a function of use-value, abundance, and degree of endangerment. Endangerment is a function of the rate at which the land cover class for a given species is being lost. Data on use patterns of the Batemi in north-central Tanzania are combined with abundance and distribution data and used to explore methodologies for identifying possible conservation priorities.

Statement of interest

Land use modeling for ecological characterization

Bradley O. Parks
University of Colorado, Cooperative Institute for Research in Environmental

John J. Kineman
National Geophysical Data Center, National Oceanic and Atmospheric

Practical methods are needed to effectively generate and communicate data/information which may constructively influence attaining and sustaining health in ecosystems subjected to increasing human pressure. First adopted in the 1970's but sparingly used since, ecological characterization is an approach to synthesizing environmental and ecological information that may be further enhanced to help meet these evolving needs. Such information syntheses can be adapted to support necessary planning and management activities and have already been conducted as intensive explanations/descriptions of ecosystems and other functional units.

A science panel convened to review thought and practice in this area has recommended that greater emphasis be given to societal, human, and behavioral aspects of ecological characterization (or "socio-ecological characterization"), both as objects of analysis and as active elements in a larger valuation and decision-making processes. An implicit consequence is that characterization will need to accommodate and articulate the complex influences of ecological processes, management objectives, human impacts, and societal values. Doing so will require integrating requisite data, information, and models within a comprehensive yet simple framework that can inform and enable evaluating options for ecosystem management.

Improved predictive capability is needed to better anticipate and direct change in both managed and unmanaged ecosystems. Land use, land transformation, and land transition modeling are key predictive elements for assessing (particularly) managed landscapes and for scenario writing methods which, when sufficiently refined, will significantly enhance the capability and robustness of modern ecological characterization.

NOAA's National Geophysical Data Center (NGDC) and the University of Colorado-CIRES through their joint Ecosystems activities are actively engaged in the development of ecological characterization as both practice and product contributing to more effective ecosystem management. Current focus is on coastal ecosystems health and coastal management for which land use modeling and its variants can become a key component. Extending land use modeling with its terrestrial emphasis through more inclusive ecosystem models to create linkages with aquatic and marine processes will provide much of the analytic capability needed to bring ecological characterization to maturity.

Consistent with the purposes described, working partnerships are being pursued by NGDC to culture both tool/technique development and methods application of land use modeling to coastal ecosystem management. The workshop provides an additional useful means of establishing such collaboration.

The Statement of Research

Hanquin Tian
The Ecosystems Center
Marine Biological Laboratory

My major interest is in the Modeling and Analysis of Land Use and Ecosystem Processes using an integrative approach including computational methods and remote sensing/GIS. Since the early 1980's, my research fields have covered a range of topics, including development and application of spatially-explicit land use change model; process-based terrestrial ecosystem modeling; development of crop production expert system; satellite image processing; landscape patterns and processes; environmental impact analysis and ecological risk assessment; the global biogeochemical cycles; etc. During 1992-1995, I worked with DOE-sponsored research group, directed by Dr. Charles A. S. Hall at SUNY, Syracuse, to develop a spatially-explicit model of land use change, called as GEOMOD. This model has been used to simulate the spatial and temporal patterns of deforestation, shifting cultivation, agricultural expansion, urbanization in tropics and their consequences to the carbon dioxide accumulation in the atmosphere and the nutrient loading in aquatic ecosystems. Two papers detailing the land use change model had been published in DOE Research Summary and Journal of Biogeography. Parts of these works have been cited by Communications, a major newsletter in global change study edited by Oak Ridge National Laboratory. On February 25, 1997, the DOE Research Summary won Technical Communication Award of the Society for Technical Communication. I was invited to chair a session about modeling complex ecological-economic system at the International Conference for Ecological Economics in August, 1996, Boston. Since September, 1995, I have been working with Dr. Jerry M. Melillo, currently on leave to OSTP, on Vegetation/Ecosystem Modeling and Analysis Project (VEMAP). As a member of VEMAP, my research is to model and analyze the inter-annual variations of terrestrial ecosystems under the changing pattern of the global climate, particularly the temperature and precipitation and atmospheric CO2 concentration. I adopted a process-based ecosystem modeling approach that enables to simulate transient changes in ecosystems. I am also involving in the research projects of NASA EOS-IDS and Max Planck Institute about terrestrial biogeochemical/ ecosystem modeling. The major limitation of current terrestrial ecosystems models is lack of explicit representative of land use/land cover change. Clearly, it is of critical importance to take into account land use/land cover change in order to better understand the global biogeochemical cycle. My current research interest includes the following four aspects:

1) Modeling land use change at regional scale
2) Modeling terrestrial ecosystem dynamics
3) Linkage of land use model with terrestrial ecosystem model
4) Parameterization, calibration and validation of ecosystem model using field and remote sensing data

Reference

1) Tian, H., H. Xu and C.A.S. Hall, 1995. Pattern and change of a boreal forest landscape in northeastern China, Water, Air and Soil Polution 82:465-476,

2) Hall, C.A.S., H. Tian, Y. Qi, G. Pontius and J. Cornell, 1995. Modeling spatial and temporal patterns of tropical land use change, J. of Biogeography, 22:753-757.

3) Hall,C.A.S., H. Tian, Y. Qi, G. Pontius, J. Cornell and J. Uhlig, 1995.Spatially explicit models of land use change and their applications to the tropics. DOE Research Summary, No. 31, edited by CDIAC.

4) Tian, H., J.M. Melillo, D.W. Kicklighter, A.D. McGuire, Y.Pan, X. Xiao and J. Helfrich. Transient responses of carbon flux and storage to increasing CO2 and temperature variability in the United States (Submitted to Ecological Applications)

Special Interests

Janet Tilley, Cartographer
U.S. Geological Survey

Special Interests

The NMD temporal mapping activity was first done by USGS in the San Francisco-Sacramento, CA., area by William Acevedo and Len Gaydos, USGS, and brought to us, per Timothy Foresman's request, to the Reston USGS to pursue temporal mapping in the Baltimore-Washington region. The next paragraph provides background information about this work and is where my interest lie. I especially would like to learn more about modeling techniques as I often must present the overall project, applications, and modeling concepts that are currently underway. I need to better understand this application. I am not a programmer, but does that mean I will not be able to do modeling?
Background

The U.S. Geological Survey, the University of Maryland Baltimore County, and the U.S. Bureau of Census have developed a temporal database to study urban development in the Baltimore-Washington region. The primary data layer, the extent of urban or built-up areas, was compiled using a geographic information system and historical maps, remotely sensed data, digital land use data, and census information from a variety of sources. Urban land use change has been documented by the Baltimore-Washington Spatial Dynamics & Human Impact Study Team for the last 200 years. The methods, definitions, and collection criteria were used to define urban or built-up areas were developed by a multi-disciplinary team that also ensured consistency in collection techniques and documentation methods for subsequent application in other regions. Animation techniques were used to visualize the database and to document the evolution of the region's urban landscape. The principal transportation data layer documents the primary roads, railroads, and other transportation features that provided the infrastructure for urban development. Compilation criteria, developed by Susan Clark, USGS, such as connectivity, mobility, lineage, and alignment were developed to accommodate limitations in the source materials. A nimating transportation features proves difficult as they are linear features and challenging to blend in and out through time, data have been refined by William Acevedo. The database is an important tool to urban and regional planners, ecologists, and g lobal change researchers for measuring trends in urban sprawl, analyzing patterns of water pollution, understanding the impacts of development on ecosystems, and developing predictive modeling techniques to better forecast areas of urban growth. Web page s created by Len Gaydos, William Acevedo, and Janis Taylor Buchanan discussing detailed information about Land Transformations, San Francisco-Sacramento, CA., and Baltimore-Washington, VA-MD-PA-WV., work are available; as well as Timothy Foresman, UMBC, web pages discusses the collaboratory and has our data that can be downloaded for further analysis.
Applications

Currently, there are several activities that are using temporal data from the San Francisco-Sacramento and Baltimore-Washington regions for applications and modeling. In applications of these types of data the Chesapeake Bay Program is using the urban temporal theme to correlate the urban growth with the change in water quality. Such issues as nutrient loading and sediment run-off are of special concern in their impact on fragile ecosystems.

William Acevedo, USGS, is testing the feasibility and generating a temporal agricultural land theme for the Patuxent watershed; and Timothy Foresman, UMBC, is testing the feasibility and generating a temporal forested land theme in the Gwynns Falls Watershed in Baltimore, MD. Both of these research endeavors are being completed for the Chesapeake Bay Program. Again the Chesapeake Bay Program will use these data to correlate with the water quality.

Modeling and Prediction

Modeling and prediction techniques of both the San Francisco-Sacramento, CA., and Baltimore-Washington, VA-MD-PA-WV., areas are being researched by Keith Clark, UCSB. One modeling technique he has applied is the application of a wild fire algorithm to that of urban sprawl. Lee De Cola, USGS, has also completed fractal modeling of the Baltimore-Washington, VA-MD-PA-WV., area, where he generated urban surface intensity animations of our data.

Future Research

This multi-discipline team is hoping to pursue the opportunity to refine their research on developing tools for generating these temporal databases, and hoping to build lasting partnerships to pursue these endeavors. Our current tools are crude and the technology could not be passed on to other groups in additional metropolotan areas. We are anticipating developing databases for the following four cites: New York, NY., Philedelphia, PA., Chicago, IL., and Portland, OR. In the anticipation of the opportunity to continue this research we are investigating source availability, and are beginning to establish contacts for additional partners and source materials in these areas.

STATEMENT OF INTEREST

MODELING OF LAND USE / LAND COVER CHANGE : A CASE STUDY IN SENEGAL

Eric C. Wood
International Program
USGS EROS Data Center

Introduction

In recent years Africa has been plagued with wars, famine, deforestation, desertification, overpopulation, and general resource depletion. There is no question that Africa is in crisis (Mortimore, 1989). Just how severe is that crisis remains the question facing the world development community as represented by such diverse groups as the United Nations, World Bank, bilateral donors, international non-governmental organizations (NGO's), and of course the countries themselves.

This "crisis" has been particularly apparent in the Sahel. Droughts and famine have long been a factor of life in this region, but mechanisms were previously in place to withstand them. Now, as population pressures, international economies, and mechanized agriculture have become more prominent, this is less the case. These factors have forced sub-Saharan Africans into land use practices that are often inappropriate for the land (Toure, 1989; Franke and Chasin, 1980; Freudenberger and Freudenberger, 1993).

It has been suggested that the increasing rate of land use/land cover change is one the of the most important ecological issues in Africa (de Graf, 1993). Because the majority of African countries have rural economies and depend heavily upon their natural resources (e.g. for food, fuelwood, commodity exports), degradation of those resources can result in rapid declines in standards of living. Senegal, in West Africa, is an example of such a country. Buffeted by drought, currency devaluation, and rapid population growth, Senegal's ecosystems are coming under increasing pressure. There is evidence of rapid changes in the landscape, which in many cases is threatening local livelihoods (Tappan and Wood, 1995).

It is also clear that degradation of natural resources in Africa cannot be treated as solely a biophysical problem (Kates, 1990; Blaikie, 1985; Bryant and LeDrew, 1989; Jacobson, 1990). For researchers in Africa, as globally, separating the human induced from the climatically induced change is a major challenge. Determining the socioeconomic factors responsible for the human induced landscape change is even more complicated and has been an evolving focus of governments and donors throughout the continent.

Objectives

The major goal of this research is to develop a methodology in the form of descriptive and predictive models, which will facilitate the determination of factors driving land use/land cover change in southern Senegal. In the process of accomplishing this, we hope to determine the most effective method(s) for modeling those changes. "Effectiveness" in this context consists of two key requirements : (1) that the models are technically straightforward enough to be applied by interested government agencies and (2) that they are accurate and flexible enough to be of use in providing realistic simulations for managers and policy makers.

The three objectives of this research are :

1. To determine the most accurate approach(es) for modeling land use/land cover change in southern Senegal. Several modeling approaches, built on the foundations of Markov processes and logistic regression and modified through implementation of rule based techniques, will be used to characterize and predict land use/land cover. Accuracy will be determined statistically based on each model's ability to characterize existing and simulate future land use/land cover. An iterative approach will be used in order develop a final model that is technically straightforward while still maintaining as high a degree of accuracy as is possible from the models being developed.

2. To determine which factors are driving land use/land cover change in southern Senegal. In order to accomplish this, land use /land cover classification and change techniques will be carried out using a historical time series of Landsat MSS and TM and ancillary data. Models will be developed to characterize land use change, as described above. From these models the significance of the contribution of each change factor will be determined.

3. To develop realistic simulations from predictive models for managers or policy makers. Upon development of the final predictive model as described above, a methodology for creating "what-if" simulations will be developed that will allow for hypothetical changes in key change factors in order to replicate the effects of a range of policy or management decisions, exogenous factors, or natural influences.

This modeling effort has the potential to assist natural resource managers, policy makers, and the scientific community in Senegal. To date there has been no extensive inventory of land use/land cover change in Senegal, let alone a systematic investigation of causal relationships or what changes are likely to occur in the future. By providing the decision making community with a technically feasible modeling approach, the opportunity exists to correct this situation. Techniques developed here may also contribute to the land use/land cover modeling knowledge base in general.

As a result of extensive field work in the study area and investigation of relevant literature and datasets, a number of research questions related to land use have surfaced. The modeling process will attempt to "answer" many of these. The following are examples of a few key questions with others becoming apparent as the research continues.

* Does the rate of population growth correspond to the rate of expansion of land into agriculture? How does the rate of population growth effect the fallow system?

* Is agricultural expansion driven more by proximity to the village or existing fields, or by appropriate soil type? To what extent is agriculture expanding into inappropriate soil types?

* Is the land use/land cover change profile in the two northern departments of the study area (Kaffrine and Tambacounda) being reflected in the two departments in the Casamance portion of the

study area (Kolda and Velingara)? How is the land use/land cover change profile effected by the stage the department is in as a charcoal source, i.e. in the pre-extraction, extraction, or post-extraction stage?

* Is expansion along major roads occurring despite inappropriate soil types? Along minor roads in inappropriate soil types? At what rates?

* Is the Niokola Koba National Park withstanding edge intrusion in areas suitable for agriculture? Does the rate of change adjacent to the park vary from the norm, i.e. is there any self-generating buffer effect? Is the same true for reserves of a different status, i.e. foret classe, sylvo-pastoral or some other non-park status?

* Does the species richness and abundance (biodiversity components) of a non-reserve forest, including the presence of "noble species", effect the rate of land use/land cover change?

* Is the previous (at time t0) land use/land cover state significant in determining the transition state (at t1)?

* What are the spatial factors (e.g. contiguity, patch size, edge, etc.) most influencing change ? How do they influence change?

* Are the activities of the agricultural parastatals (e.g. Sodifitex, Sodagri) accelerating the rate of change?

* What influence have key policy reforms (e.g. agriculture, forestry, land laws) had on the rates and types of land use / land cover change?

Study site

The general study area consists of four departments located in south-central Senegal (Kaffrine, Tambacounda, Kolda and Velingara). There are 30 departments in all of Senegal, the four above being of particular interest in that they have abundant natural resources, fall within the zone of viable rainfed agriculture, and as a result are undergoing significant land use/land cover change. The initial investigation will be carried out in the department of Velingara.

This department, like the other three, faces a number of critical land use issues, including the effects of charcoal and fuelwood production, cultivation of cash crops (peanuts and cotton), excessive burning, and agricultural expansion.

Data

The primary focus for the socioeconomic data collection was to determine land use management practices and general resource management efforts, but has resulted in collection of a variety of other data. These consist of site observations, intermittent field contacts, short village visits to answer specific field questions, and more in-depth structured village studies. The latter have been carried out in five ecologically distinct villages with the use of rapid rural appraisal (RRA) techniques (e.g.

Freudenberger, 1996) . Also, numerous half-day village interviews have been carried out using some of these same RRA techniques, most notably the semi-structured interview.

Other socioeconomic datasets include USAID's three Knowledge, Attitudes, and Practices (KAP) studies; the results of the RRAs carried out in USAID's seven village REMAP study ; village specific RRAs by other investigators (e.g. Freudenberger and Freudenberger, 1992); and numerous project reports by the Government of Senegal and various bilateral or multilateral donors .

Modeling

A series of models of increasing complexity will be developed in order to determine the range of accuracy that can be obtained. In each case the model will be calibrated and validated using the change series generated from the classification of the 1973, 1978, 1985 and 1990 Landsat MSS scenes for the department of Velingara. The models will then be used to predict the land use/land cover in 1995 from 1990 or previous data and subsequent accuracy determined. The best performing models will then be applied to another department (within the general study area - refer to Section 4.1) in order to determine their general robustness. The choice of 1995 is based on the approximate 5 year time interval between historical images

The four general categories of models to be investigated consist of :

Simple Markov
Markov model with spatial modifier
Logistic regression
Knowledge based system

Results

It is anticipated that the results of this research will include:

1. Spatially explicit biophysically and socioeconomically driven models developed to describe existing landscape change . Output will include maps, GIS coverages, and tabular datasets explicitly designating the spatial extent of both actual and modeled changes in land cover.

2. Comparative analysis of the above models in terms of descriptive and predictive accuracy.

3. A qualitative assessment of the data processing and development expenditures required for implementation of each model.

4. A design for implementation of the above models as a tools for applying hypothetical changes in key change factors in order to replicate the effects of a range of policy or management decisions, exogenous factors, or natural influences.

5. In depth discussion of the relationship between biophysical and socioeconomic factors and land use/land cover change; based both on statistical evidence as well as, on convergence of evidence (i.e. accommodates RRA, etc., that may not be able to be captured in a model of types developed).

REFERENCES

Blaikie, P. (1985). The Political Economy of Soil Erosion in Developing Countries. New York: Longman Scientific & Technical.

Bryant, R.B., and E.F. LeDrew (1989). The Scope of Land Change Research. In R. B. Bryant,E.F. LeDrew,C. Marios,and F. Cavayas (Eds.), Remote Sensing Methodologies of Land Use Change Analysis. Waterloo, Ontario: University of Waterloo.

de Graff, J. (1993). Soil Conservation and Sustainable Land Use. Amsterdam: Royal Tropical Institute.

Franke, R.W., and B.H. Chasin (1980). Seeds of Famine: Ecological Destruction and the Development Dilemma in the West African Sahel. New York: Universe Books.

Freudenberger, K.S., and M.S. Freudenberger (1992). Fields Fallow, and Flexibility : Natural Resource Management in Ndam Mor Fademba, Senegal. (RRA Report). World Vision.

Freudenberger, K.S. (1996). Kandia: Threatened Forest, Threatened Livelihood. Report, USAID/Dakar. (In press).

Jacobson, H.K., and M.F. Price (1990). A Framework for Research on the Human Dimensions of Global Environmental Change. ISSC Standing Committee on the Human Dimensions of Global Change.

Kates, R.W., W.C. Clark, V. Norberg-Bohm, and B.L. Turner II (1990). Human Sources of Global Change : Priority Research Initiatives for 1990-1995. Providence: Institute for International Studies.

Mortimore, M. (1989). Adapting to Drought : Farmers, Famines and Desertification in West Africa. Cambridge: Cambridge University Press.

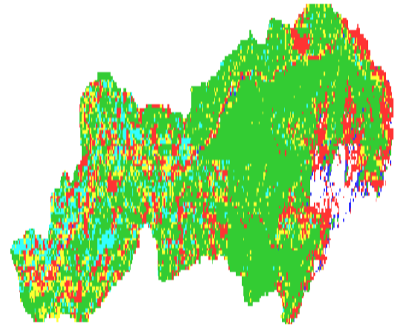
Tappan, G.G. and E.C.Wood (1995).Structure pour le suivi a Long Terme des Ressources Naturelles du Senegal. Rapport Final.Colloque National sur La Gestion Des Ressources Naturelles Au Senegal.

Toure, O. (1991). Where Herders Don't Herd Anymore: Experience from the Ferlo, Northern Senegal. (Internal Report).Dakar, Senegal: Centre de Suivi Ecologique.

Full Papers

Distributed Land-Cover Change Simulation Using PVM and MPI

M.W. Berry and K.S. Minser
Department of Computer Science
University of Tennessee
Knoxville, TN, 37996-1301



This research has been supported by the **Southeastern Appalachian Man and the Biosphere (SAMAB) Program** under **U.S. State Department** Grant No. 1753-000574 and **University of Washington** Subcontract No. 392654, by the **National Science Foundation** under grants NSF-ASC-94-11394 and NSF-CDA-95-29459, and the **USDA Forest Service** under Contract No. 29-1286.

Abstract:

Computer simulations are used in landscape ecology to simulate the effects of human land-use decisions on the environment. Such decisions are influenced by both ecological and socioeconomic factors which can be represented by spatially explicit multidisciplinary data. The **L**and-**U**se **C**hange **A**nalysis **S**ystem (or**LUCAS**) was developed to study the effects of land-use on landscape structure in such areas as the Little Tennessee River Basin in western North Carolina and the Olympic Peninsula of Washington state. These effects include land-cover change and species habitat suitability. Using a geographic information system (GIS) to store, display and analyze map layers derived from remotely sensed images, census and ownership maps, topological maps, and output from econometric models, a parallel/distributed version of LUCAS (pLUCAS) was developed for simulations on a network of workstations. Targeting distributed computational environments reflects the resources available to most land-use planners, forestry personnel, and wildlife managers. A performance evaluation of two pLUCAS distributed models on an ATM-based network of 12 SUN Ultra-2 workstations is presented. Particular emphasis is given to the range of speed improvements (relative to serial runs on a single SUN Ultra-2 workstation) that can be obtained using the PVM or MPI message-passing environments.

1. Introduction

Humans can have a direct influence on changes in the natural environment. One approach toward a better understanding of the the effects of human land-use decisions on the environment is to consider both ecological and socioeconomic factors. Such a multidisciplinary approach was taken by the Man and the Biosphere (MAB) project [2], whose goal was to analyze the environmental consequences of alternative land-use management scenarios in two different geographic regions: the Little Tennessee River Basin (LTRB) in North Carolina and the Olympic Peninsula in Washington State.

The MAB approach involved the integration of disciplines such as ecology, economics, sociology, and computer science to evaluate the impacts of land-use. This integration also required that data from the various disciplines share a compatible representation. Such forms include tabular and spatial databases, results of mathematical models, spatial models and expert opinions [2, 7]. A *geographic information system* or GIS, such as the Geographic Resources Analysis Support System (GRASS) [12], can be used to easily store and manipulate the spatially explicit representation of this data. The Land-Use Change Analysis System (LUCAS) is a prototype computer application specifically designed to integrate the multidisciplinary data stored in GRASS and to simulate the land-use policies prescribed by the integration model.

1.1 Sample Scenario and Validation

In LUCAS, scenarios describe prescribed land-use policies to be simulated. As an example, suppose that a natural resource manager in the LTRB would like to determine the impact of not logging any trees for 50 years on the habitat of the Wood Thrush (*Hylocichla mustelina*). The scenario is formally defined to use the historical transition probabilities based on existing map layers from 1975, 1980 and 1986 along with the restriction that once a grid cell of land is forested, it will remain forested. For example, the land manager may run LUCAS with 10 replicates for 10 time steps each to simulate the change over 50 years. The manager can examine the graphical statistics plotted on the screen or more carefully analyze the statistics saved to a SAS [8] file. Other scenarios with different constraints can be investigated and their results compared. In this way, the investigator can better understand the effects of potential land-use decisions.

To validate the LUCAS model, historical data are compared against the simulated data [2, 7]. Starting with the oldest existing map, the period of time up to the year for which the newest map exists must be simulated. The degree to which the statistics for the simulated and historical land cover layers agree determines the accuracy of the model for this period.

1.2 LUCAS/pLUCAS Development

The initial LUCAS prototype was implemented as an *object-oriented* C++ application to promote modularity. This modularity facilitates the addition of future software which might address the needs of different types of users. Future expansions of LUCAS are discussed in [2] and [7] while Section 2 describes (in brief detail) the modular implementation of the initial LUCAS prototype followed by the more recent parallel and distributed versions in Section 3. The creation of a distributed version [6] of LUCAS, Parallel LUCAS (pLUCAS), was motivated by the computational needs of real-time processing and extensions to larger regions. The first design of pLUCAS [6] utilized the Parallel Virtual Machine (PVM) [5] message-passing environment, and the (current) follow-up implementations are based on MPI [9]. The performance of both PVM and MPI implementations when tested on an ATM-connected network of workstations will be discussed in Section 4.

2. LUCAS Design

As discussed in [2, 7], LUCAS provides a stochastic model for the future assessment of landscape change using historical maps of land cover. The initial design's modularity provides great flexibility for future modifications required by diverse users.

2.1 Stochastic Modeling

The econometric model used in LUCAS is a dynamic, *stochastic* model primarily based on one random variable, namely land cover, and deals explicitly with time-variable interaction. The stochastic simulations enabled by LUCAS employ the statistical sampling of multiple replicates, i.e., repeated simulations of the same model. The statistical output produced by LUCAS is composed of SAS-compatible [8] data which can be imported by any generic graphing tool/software. Figure 1 outlines the modular model used to develop the LUCAS prototype. Each module of the LUCAS model is briefly described in the following sections (see [2] for more details).

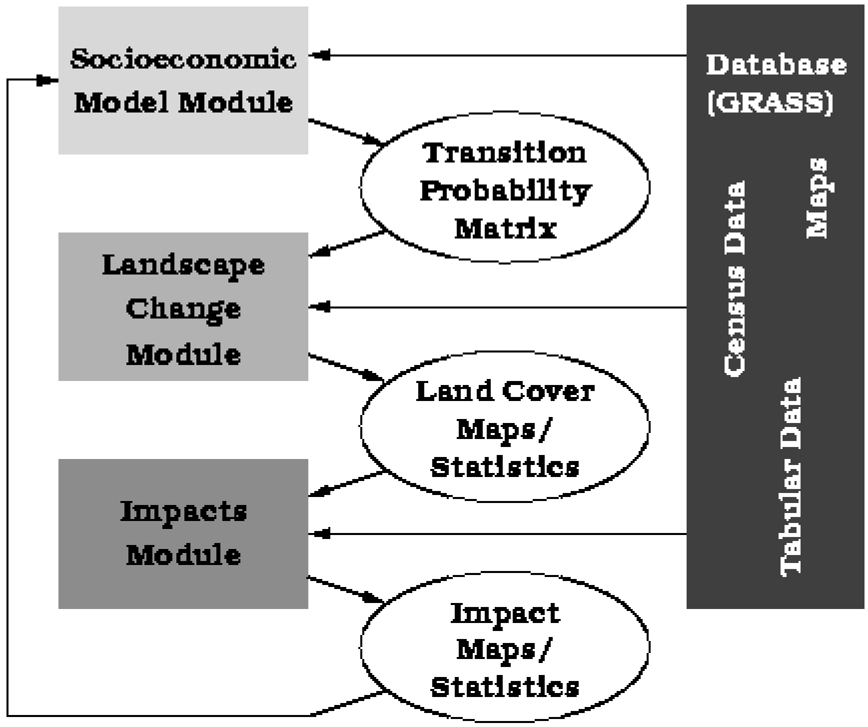


Figure 1: Relationship among LUCAS modules

2.2 Socioeconomic Model Module and TPM

Several discrete and continuous ecological and sociological variables were used empirically in calculating the probability of change in land cover: land-cover type (vegetation), slope, aspect, elevation, land ownership, population density, distance to nearest road, distance to nearest economic market center (town), and the age of trees. For an analysis of the influence of these economic and environmental factors on landscape change see [11]. Each variable corresponds to a spatially explicit map layer stored in the GIS. A vector of all of these values for a given grid cell is called the **landscape condition label**

[3, 4]. An example landscape condition label (LCL) [7] is shown in Table 1.

ℓ	x_ℓ	Definition	Attribute
1	2	Public Lands	Land ownership
2	75	75 years old	Tree age (Olympic Peninsula only)
3	1	Coniferous	Land cover (vegetation)
4	20	20° incline	Slope
5	1	True	Steep slope (> 17° incline; Olympics only)
6	1500	1500 meters	Elevation
7	1000	1000 meters	Distance along roads to nearest market center
8	21	1890 meters	Distance to nearest road
9	2	0.002 people/acre	Population density (LTRB only)

Table 1: Example Landscape Condition Label in the Hoh Watershed on the Olympic Peninsula

Each element of the LCL $\vec{x} = (x_1, x_2, \ldots, x_8)^T$ is used to determine the probability of change using the multinomial logit equation [14, 13, 2]

$$\Pr(i \rightarrow j) = \frac{\exp(\alpha_{i,j} + \vec{x}^T \vec{\beta}_j)}{1 + \sum_{k \neq i}^{n} \exp(\alpha_{i,k} + \vec{x}^T \vec{\beta}_k)}, \qquad (1)$$

where n is the number of cover types, $\vec{x}$ is a 5×1 column vector composed of elements $x_4, \ldots, x_8$ of the LCL $\vec{x}$ in Table 1, $\vec{\beta}_j = (\beta_{1,j}, \beta_{2,j}, \ldots, \beta_{5,j})^T$ is a vector of logit coefficients, $\alpha_{i,j}$ is a scalar intercept, and Pr $(i \rightarrow j)$ is the probability of coniferous land cover remaining the same ($i = x_3 = 1 = j$) at time t+1 or changing to another cover class (i.e., j=2,3,4). The land ownership (x_1) determines which table of logit coefficients should be used and the tree age (x_2), used only for coniferous forest cover types, determines if the trees have aged sufficiently to be harvested, i.e., change to another cover type. The **null-transition** or probability of no land cover change is defined by Equation 2.

$$\Pr(i \rightarrow i) = \frac{1}{1 + \sum_{k \neq i}^{n} \exp(\alpha_{i,k} + \vec{x}^T \vec{\beta}_k)}, \qquad (2)$$

where the symbols have the same meaning as in Equation (1). Example vectors of coefficients for the Hoh and LTRB Watersheds are available in [2] and [7]. Such coefficients and associated intercept values have been calculated empirically by Wear *et al.* [14] from existing historical data stored in the GRASS database. The table of all probabilities generated by applying Equation (1) to all cover types is called the **transition probability matrix** (TPM), an example of which can be found in Table 2. If the TPM in Table 2 were used, for example, a random number from the closed interval [0,1] less than 0.8725 would signify that the land cover would remain coniferous. For a discussion of logistic regression and a basis for Equation (1) see [10].

From Coniferous	Changing to	Probability
$1 \rightarrow 1$	Coniferous	0.8725
$1 \rightarrow 2$	Deciduous/Mixed	0.1186
$1 \rightarrow 3$	Grassy/Brushy	1.886×10^{-3}
$1 \rightarrow 4$	Unvegetated	6.989×10^{-3}

Table 2: Example Transition Probability Matrix based on the example multinomial logit coefficients.

2.3 Landscape Change Module

The Landscape Change Module in Figure 1 is the heart of the LUCAS software. On input, this module accepts the multinomial logit coefficients generated in Socioeconomic Model Module, implements the actual landscape change, and produces new land cover maps and statistics as output. The first step in designing LUCAS was to develop the method to simulate one time step, a five year period, of landscape change over multiple replicates.

Two types of transitions are simulated by LUCAS: grid cell (or pixel-based) and patch-based. The determination of the pixel-based landscape transitions is relatively trivial because each grid cell changes independently. The transition probabilities from the initial cover type to all other cover types are calculated using Equation (1) and the value of the landscape condition label of a grid cell. A pseudorandom number is then drawn from a uniform distribution between 0 and 1. This number, in turn, determines the new land cover type for this grid cell via the calculated probabilities. Patch-based transitions are considerably more difficult because of the task of patch identification. A patch (or cluster) is a group of contiguous grid cells with identical landscape condition labels. Although patch identification was not used in this research effort, algorithms for determining both the number and structure of patches (clusters) is available [2].

Once the map of new land cover has been generated, the ecologist or land manager can use the results to determine the impact of the policy defined in the Socioeconomic Model Module. As stated in Section 2.1, statistics are the only true metric for analyzing a stochastic simulation. They also provide a convenient method for understanding the impact of the particular land management policy or scenario. The statistics in Table 3 are collected by LUCAS for each time step.

	Statistic
Pixel Statistics	Proportion of landscape in each cover type Area (ha) of landscape in each cover type Edge:area ratio for each cover type Amount of edge (km) for each pair of cover types Total edge (km) in the whole landscape
Patch Statistics	Number of patches Mean patch size Standard deviation of patch size Size of largest patch Cumulative frequency distribution of number of patches by size Mean patch shape (normalized shape index)[1]

Table 3: Statistics collected by LUCAS

2.4 Impacts Module

The land cover maps produced by the Landscape Change Module (see Section 2.3) are analyzed by the Impacts Module. This module may eventually determine the effect the changed landscape has on species, habitats, water quality, or other environmental impacts. Currently LUCAS is designed to perform only species' habitat suitability analyses [2, 7]. Although an extensive list of species and habitat identification algorithms for each of the watersheds currently simulated are available, this module was not used in the results presented in Section 4. The usual output from this particular module is a binary map; either a grid cell is suitable for a species or it is not. The statistics in Table 3 are again collected for each impact map.

3. pLUCAS Implementations Using PVM and MPI

The parallel and distributed implementation of LUCAS (pLUCAS) is based on the same functional design of the serial prototype described in Section 2. The motivation for pLUCAS was to manage the multiple independent replicates required (for accuracy purposes) in the stochastic simulation of land-cover change. As most end users of LUCAS would not be expected to have multiprocessor computing systems available, software that could exploit a network of workstations was considered more desirable.

Parallelization is used so that each processor performs one complete simulation (replicate) of LUCAS exclusive of any other processor or process. The statistics calculated from each replicate are stored locally to disk (on each processor) until all replications are completed. At that time, the statistics are assembled on the main node and stored for later use. In the performance tests presented in Section 4, 10 replicates of 20 timesteps each are performed for 4 different scenarios on the Hoh Watershed of the Olympic Peninsula. Along with these 40 tasks (4 scenarios $\times$ 10 replications) is a task for the initialization of any future impact modules that could be used (see Section 2.4) so that a total of 41 independent tasks are scheduled.

The initial version of pLUCAS was implemented using PVM [6] and has now been modified to use the MPI message-passing software library [9]. All versions emulate the basic host/worker model (described below) with some differences inherent to PVM and MPI. All pLUCAS runs using PVM and MPI were tested on a network of twelve Sun Microsystems Ultra 2 workstations, each containing two 167-Mhz UltraSPARC-1 processors under the Solaris 2.5.1 operating system. Each workstation had 256-Mbytes of memory and two 2.1-Gbyte internal disks. Peak performance of one UltraSPARC-1 processor is about 126 Mflops (millions of floating-point operations per second). The workstations were connected by both a 10 Mbps Ethernet interface and 155 Mbps ATM sbus adapter so that performance results (recorded in elapsed wall-clock time) could be obtained with two different network latencies.

3.1 PVM Version

The PVM (version 3.3.10) implementation of pLUCAS allows the host process to assign tasks to worker processors by spawning a worker process onto a specific machine. The host maintains a queue of tasks to be scheduled, tasks completed, and available workers (machines). After all tasks are completed, the host spawns new processes on all machines to send data accumulated from their previous tasks back to the machine owning the host process. The host process collects the data, assembles it, and writes the results to files stored on a machine external to the ATM-connected network. The host processor is allowed to spawn a worker process to itself. Thus, the host machine will have one host and one worker process assigned to it. All other machines will have only one worker process at a time.

3.2 MPI Versions

To incorporate MPI (version 1.0.13) into the pLUCAS software, major code revisions were necessary due to the lack of process spawning with MPI. A traditional **host/worker** model was implemented using a top-level if-then-else construct to select the appropriate set of instructions for each process type. All tasks/duties assigned to the workers and the final accumulation of statistics from all processes (see Section 3.1) are accomplished via message-passing.

Two different versions of MPI have been developed. The first method does not allow a worker process to coexist on the same machine as the host process. Therefore, a 4-machine network has only 3 worker processes as shown in Figure 2(a). This setup is referred to as MPI(1) in all subsequent tables and graphs and is defined as k processors on k machines. The second MPI version is a better emulation of the PVM approach (see Section 3.1) which allows for 1 worker process to be assigned to the host process machine. Therefore, a 4-machine network would have one host process and 4 worker processes. This method is referred to as MPI(2) on all subsequent tables and graphs and defined as *k* processes on *k*-1 machines (see Figure 2(b)).

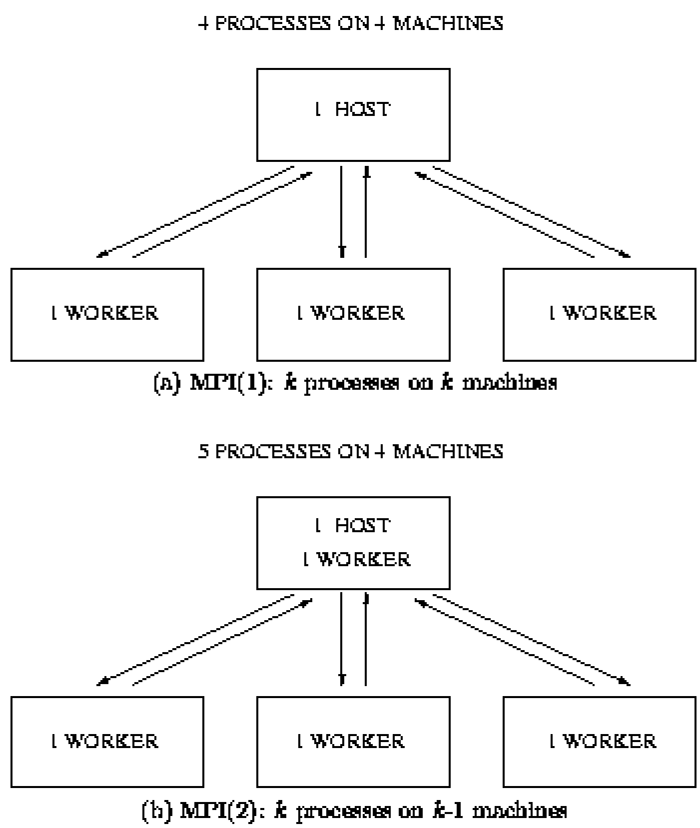


Figure 2: Two MPI-based implementations of the distributed pLUCAS model.

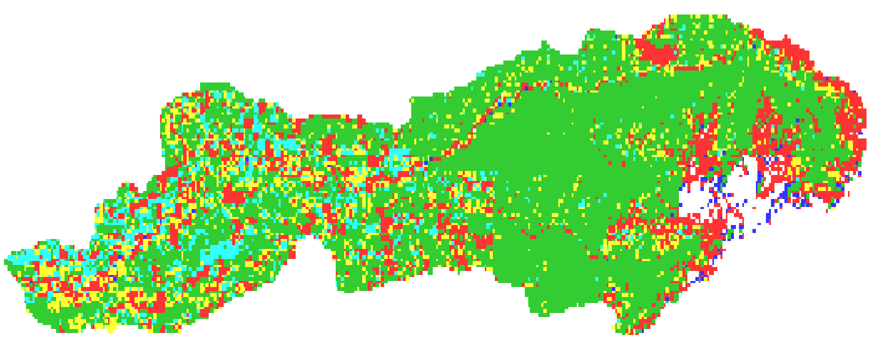
4. Performance Results for pLUCAS

In order to test the speed and scalability of the pLUCAS implementations described above, several experiments were conducted on the network of Sun Ultra 2 workstations described in Section 3. These experiments involved 3 runs each with 1, 2, 4, 8, and 12 workstations to compute 10 replicates of 20 time steps for each of the four historical, pixel-based scenarios of the Hoh Watershed on the Olympic Peninsula shown in Table 4. Figure 3 illustrates one of the Hoh land-cover maps obtained before and after a 100-year simulation (using Scenario 1 from Table 4).

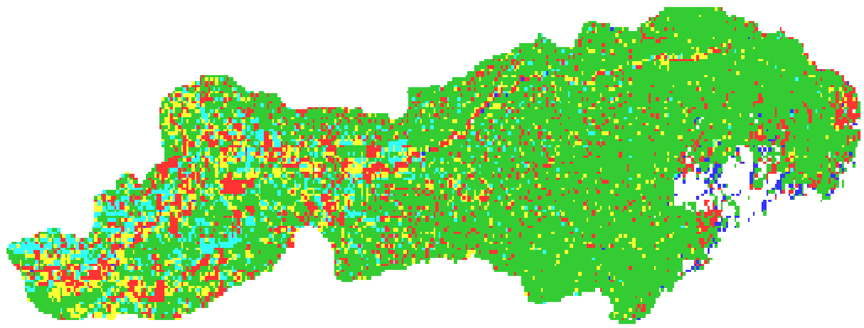
Results reported for a single Sun Ultra 2 workstation reflect the use of the serial LUCAS implementation (i.e., no message-passing overhead). The elapsed wall-clock times recorded for the experiments reported in this section are provided in Tables 5 and 6 in the **Appendix.** These wall-clock times do not include the installation of GIS data (done only once before any experiments were run) on a local disk of each machine on the ATM-connected network.

Scenario	Ownership Type	
	Public	Private
1	1986–1991	1986–1991
2	1986–1991	1975–1986
3	1975–1986	1986–1991
4	1975–1986	1975–1986

Table 4: Scenarios of land-cover change for Hoh Watershed according to historical transition probabilities



(a) Before simulation of Scenario 1



(b) After simulation of Scenario 1

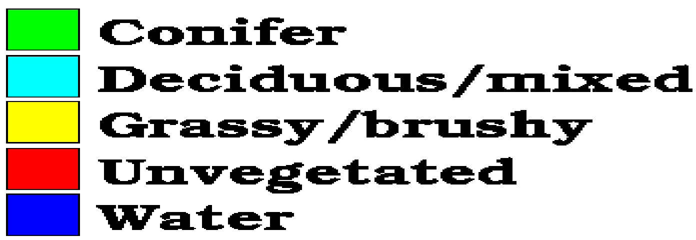


Figure 3: Hoh Watershed maps before and after a 100-year simulation

4.1 PVM Versus MPI

For both PVM and MPI, the ATM (155 Mbps) network was slightly faster than the Ethernet (10 Mbps) network, but the difference was not significant (see Figures 4 and 5). The fact that the only message-passing in pLUCAS occurs between the host and the workers (none between workers) might account for the small improvement of ATM over Ethernet. PVM had the best performance over either method of MPI with MPI(2) performing most similar to PVM. However, as more machines are added to the network, MPI(2) began losing its advantage over MPI(1) (see Figure 6).

Serial time for LUCAS on a single machine was 28.27 minutes. With 12 machines, the PVM implementation completed in 3.61 minutes yielding a speedup of 7.83. For the same number of machines, MPI(1) finished in 3.86 minutes followed by MPI(2) in 3.87 minutes with speedups of 7.32 and 7.30, respectively. With a 4-node machine, MPI(2) out-performed MPI(1) with speedups of 3.29 and 2.68, respectively. For an 8-node machine, MPI(2) yielded a speedup of 5.54, but MPI(1) maintained a speedup of 5.47. Finally, increasing the machine to 12 nodes resulted in MPI(1) and MPI(2) showing similar performances with MPI(1) actually being slightly faster than MPI(2).

4.2 MPI(1) Versus MPI(2)

The faster deteriorating performance of MPI(2) compared to that of MPI(1) can be attributed to process contention. Recall that for the MPI(1) scheme, the host process is scheduled on a dedicated processor and does not compete with any worker process. As more machines are included in the network, more message-passing demands are required of the host. In the MPI(2) scheme, the host machine has a host process as well as a worker process competing for the same port for message-passing. Note that although each Sun Ultra 2 workstation had two 167-Mhz UltraSPARC-1 processors, contention for the single ATM port degraded message-passing latencies when more than 1 PVM or MPI process was scheduled on a given machine. As the number of messages increase with added machines, the network contention becomes aggravated for MPI(2).

In order to validate the network contention suffered by MPI(2), the idle time or *wait-time* for message-passing was measured on a 12-node machine. The elapsed wall-clock time for waiting was measured using the function **MPI_WTime()** which returns the actual wall-clock time in seconds. Two calls to**MPI_WTime()** were made (before and after each blocking send and receive). The time returned from the first call is subtracted from the second call to yield the elapsed time of the message-passing function. The wait-times of the message-passing routines were summed for an entire run on each processor. The accumulated wait-time incurred is divided by the number of tasks assigned to that particular processor to determine the average wait-time per task on each processor (workstation). In some circumstances, not all processors performed the exact same number of tasks within a complete experiment. These averages were determined from 6 runs of both MPI(1) and MPI(2). The range of wait-times as well as the mean are illustrated in Figure 7 and listed in Table 6 of the Appendix.

For both MPI(1) and MPI(2), processor 3 contained the host process, and in the case of MPI(2), processor 3 contained both the host and a worker process. Although the wait-time of processor 3 for MPI(2) was not significantly high, the wait-times for MPI(2) across all 12 processors was much higher than those obtained with MPI(1) inferring a more congested network for MPI(2). Adding more processors yielded greater wait-times, and hence the time improvements for the MPI(2) quickly deteriorated.

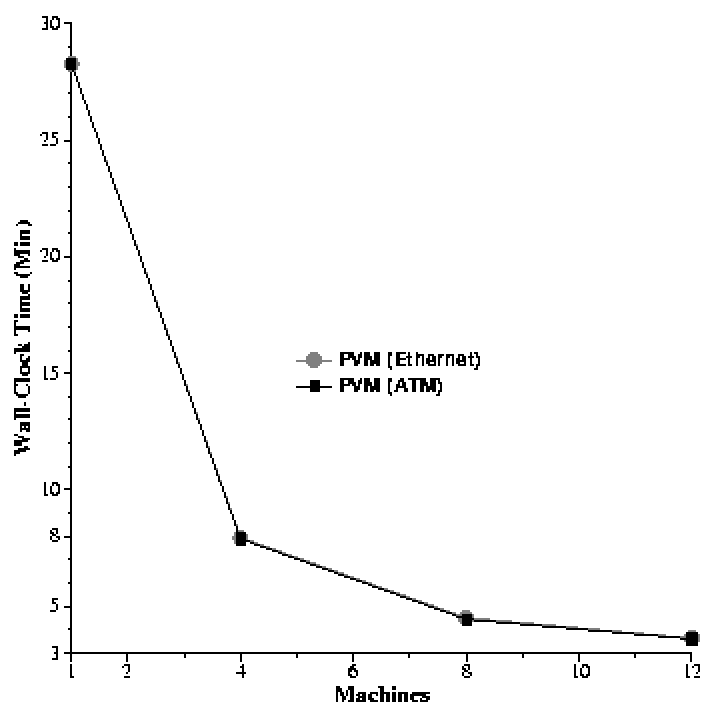
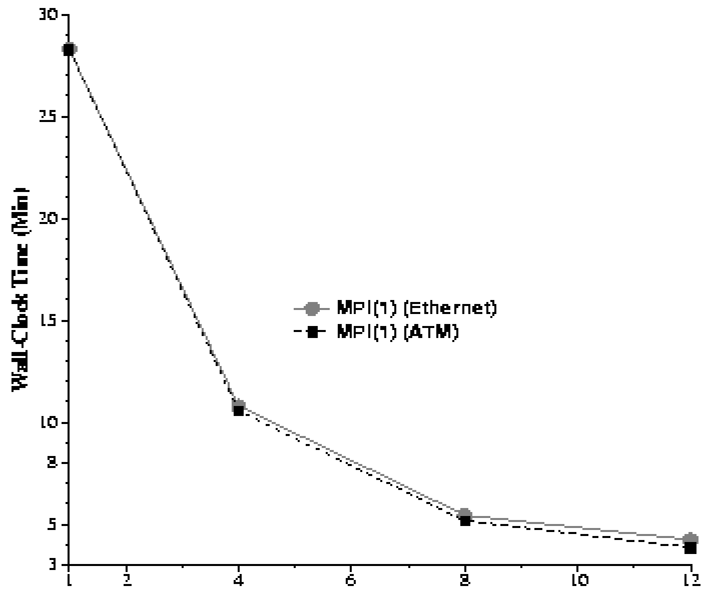
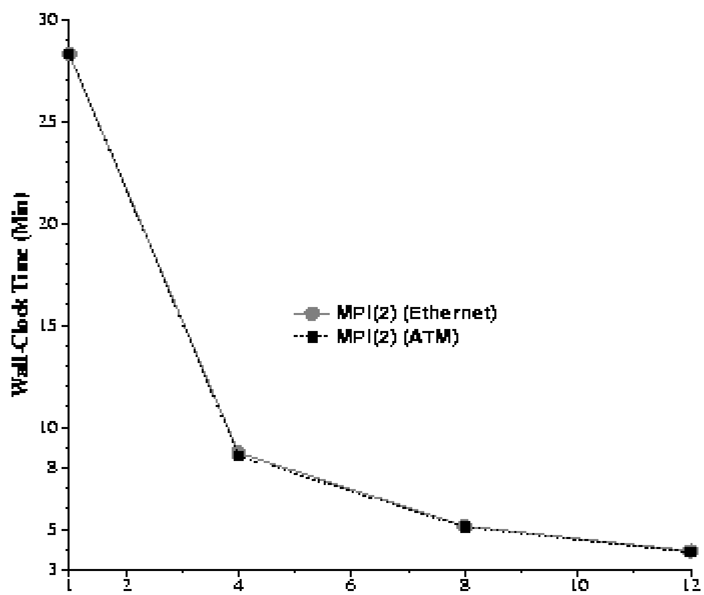


Figure 4: Timing comparisons for PVM-based implementation on ATM and Ethernet.



(a) MPI(1) performance (k processes on k machines)



(b) MPI(2) performance (k processes on k-1 machines)

Figure 5: Timing comparisons for MPI(1,2) implementations of pLUCAS using ATM and Ethernet connections.

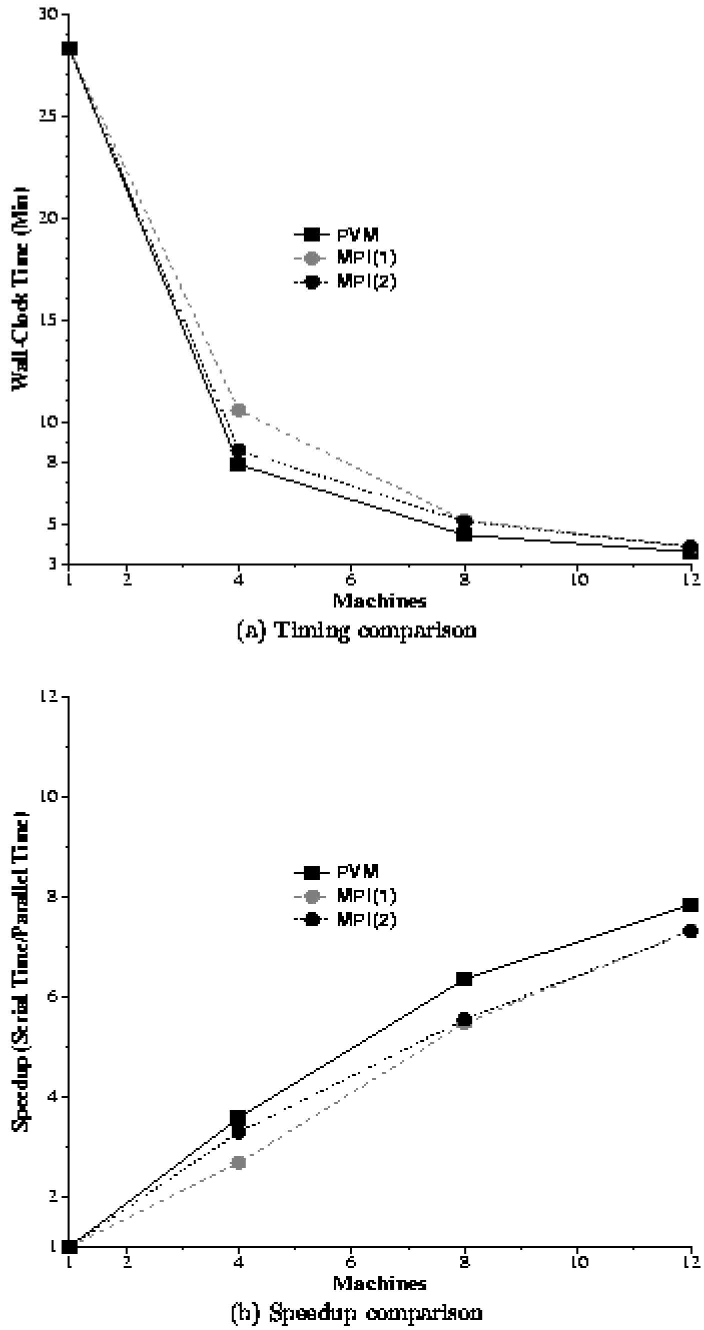
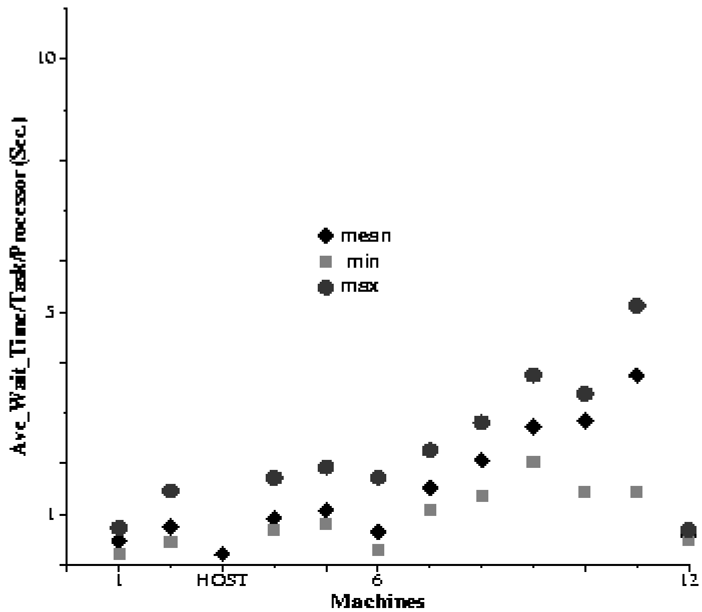
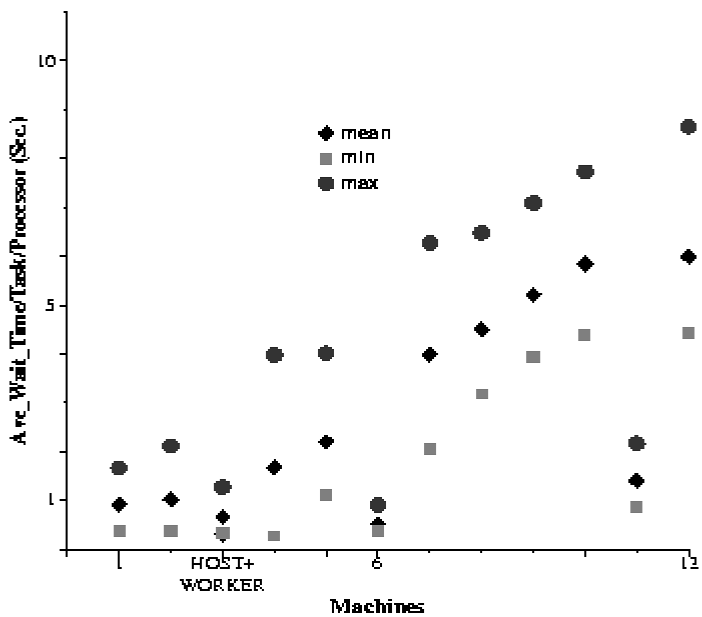


Figure 6: Timing and speedup comparisons of all three pLUCAS versions (PVM, MPI(1), and MPI(2)) using an ATM-connected network.



(a) MPI(1) implementation (k processes on k machines)



(b) MPI(2) implementation (k processes on k-1 machines)

Figure 7: Average wait-times per process for each machine when using the MPI(1,2) implementations of pLUCAS on an ATM-connected network.

5. Conclusions and Future Work

The Land-Use Change Analysis System (LUCAS) is a valuable problem solving environment for modeling landscape changes. pLUCAS offers a distributed solution to computational demands of stochastic simulation on a network of workstations. No significant differences were observed in the performance of ATM versus Ethernet with the PVM and MPI implementations of pLUCAS. Although the PVM implementation did produce faster execution times for all numbers of machines on the network, the MPI(2) implementation did perform equally well. As the network size grew, the network congestion suffered by MPI(2) offset the potential speedup gain with more machines. The host/worker distributed model used in MPI(1) was certainly less sensitive to aggravated network congestion since the host process did not share machine resources with any worker process.

Future software development of the pLUCAS prototype includes the porting of the MPI implementations to a recently acquired IBM SP-2 multiprocessor system (having 40 computational nodes), and a more thorough investigation of how two or more processes scheduled on one of the Sun Ultra 2 workstations (having two physical processors) can better time-share a single ATM port. Future modeling work with LUCAS and pLUCAS includes the integration of multidisciplinary data from several forestry growth and production models (funded by the Environmental Impacts Program of the USDA Forest Service).

References

1

W. L. Baker and Y. Cai. The r.le programs for multiscale analysis of landscape structure using the GRASS geographical information system. Landscape Ecology, 7(4):291-302, 1992.

2

M. W. Berry, R. O. Flamm, B. C. Hazen, and R. L. MacIntyre. Lucas: A System for Modeling Land-Use Change. IEEE Computational Science and Engineering, 3(1):24-35, June 1996.

3

R. O. Flamm and M. G. Turner. Alternative model formulations for a stochastic simulation of landscape change. Landscape Ecology, 9(1):37-46, 1994.

4

R. O. Flamm and M. G. Turner. GIS applications perspective: Multidisciplinary modeling and GIS for landscape management. In V. Alaric Sample, editor, Forest Ecosystem Management at the Landscape Level: The Role of Remote Sensing and Integrated GIS in Resource Management Planning, Analysis and Decision Making, pages 201-212. Island Press, Washington, D.C., 1994.

5

A. Geist et al. PVM: Parallel Virtual Machine A User's Guide and Tutorial for Networked Parallel Computing. MIT Press, Cambridge, MA, 1994.

6

B. C. Hazen. A Distributed Implementation of the Land-Use Change Analysis System (LUCAS) Using PVM. Master's thesis, University of Tennessee, Knoxville, August 1995.

7

B. C. Hazen and M. W. Berry. The Simulation of Land-Cover Change Using a Distributed Computing Environment. Simulation Practice and Theory, 1997. In Press.

8

SAS Institute Inc., Cary, NC. SAS User's Guide: Statistics, 1992.

9
M. Snir, S. Otto, S. Huss-Lederman, D. Walker, and J. Dongarra. MPI: The Complete Reference. The MIT Press, Cambridge, MA, first edition, 1996.

10
J. C. Trexler and J. Travis. Nontraditional regression analyses. Ecology, 74(6):1629-1637, 1993.

11
M. G. Turner, D. N. Wear, and R. O. Flamm. Influence of Land Ownership on Land-Cover Change in the Southern Appalachian Highlands and the Olympic Peninsula. Ecological Applications, 1996. In Press.

12
U.S. Army Construction Engineering Research Laboratory, Champaign, IL. GRASS 4.1 User's Reference Manual, Spring 1993.

13
D. N. Wear and R. O. Flamm. Public and private forest disturbance regimes in the southern Appalachians. Natural Resource Modeling, 7(4):379-397, Fall 1993.

14
D. N. Wear, M. G. Turner, and R. O. Flamm. Ecosystem management in a multi-ownership setting: Exploring landscape dynamics in a southern Appalachian watershed. Ecological Applications, 1995. In Press.

7. Appendix

(a) Wall-clock times in minutes

	PVM		MPI(1)		MPI(2)	
Machines	Ethernet	ATM	Ethernet	ATM	Ethernet	ATM
4	7.96	7.90	10.80	10.56	8.75	8.60
8	4.52	4.45	5.44	5.17	5.15	5.10
12	3.64	3.61	4.24	3.86	3.93	3.87

(b) Speedups

	PVM		MPI(1)		MPI(2)	
Machines	Ethernet	ATM	Ethernet	ATM	Ethernet	ATM
4	3.55	3.58	2.62	2.68	3.23	3.29
8	6.25	6.35	5.20	5.47	5.49	5.54
12	7.77	7.83	6.67	7.32	7.19	7.30

Table 5: Wall-clock times (in minutes) and speedups for all three pLUCAS implementations using Ethernet- and ATM-connected networks.

(a) MPI(1)

nprocs	1	2	3 (H)	4	5	6	7	8	9	10	11	12
min	0.23	0.46	0.21	0.67	0.79	0.28	1.10	1.37	2.05	1.43	1.43	0.51
mean	0.47	0.75	0.21	0.92	1.07	0.65	1.52	2.07	2.72	2.84	3.74	0.57
max	0.72	1.46	0.21	1.72	1.93	1.73	2.26	2.81	3.75	3.38	5.12	0.68

(b) MPI(2)

nprocs	1	2	3 (H)	3 (W)	4	5	6	7	8	9	10	11	12
min	0.36	0.37	0.24	0.35	0.27	1.11	0.37	2.06	3.18	3.94	4.38	0.88	4.43
mean	0.91	1.02	0.30	0.66	1.68	2.20	0.52	3.99	4.50	5.21	5.84	1.40	5.99
max	1.66	2.11	0.36	1.27	3.98	4.01	0.91	6.27	6.48	7.09	7.73	2.17	8.65

Table 6: Wait-time per task per machine for MPI(1,2) on each machine using an ATM-connected network.

Land Transition Modeling With Deltatrons

Keith C. Clarke
Department of Geography
University of California, Santa Barbara
Santa Barbara
CA 93106

Abstract

Following on from extensive work on simulating urban growth with cellular automata, a model has been developed to extend the influence of the urban growth into Anderson Level I Land Use/Cover transitions. We have built a model using cellular automata working as artificial agents in the transition phase of land cover change. Called Deltatrons, these "bringers of change" enforce spatial and temporal autocorrelation of the land cover transition process. We are now testing the limits of the model, and hope to apply it both to our San Francisco data and to the entire lower 48 United States.

Introduction

In previous work, we have examined extensively the transitions that take place as an urban area expands. Early work built a land transition model for only two land use/cover states (urban and non-urban), which was calibrated using historical data for the San Francisco Bay region. The model was based on a cellular automaton, with its spatial behavior rules modified to reflect topography, transportation, and land exclusion. After extensive calibration, the model was used for prediction over long time periods. The work has since been extended with a second major application to the Washington/Baltimore metropolitan area. Calibration results for the two applications are reported in a recent paper (Clarke and Gaydos, 1997). In addition, a localized application to the Sterling Forest area of New York allowed future scenario planning and GIS integration for decision making (Kramer, 1996).

Our latest work has extended the scope of the modeling component of the work to land transitions. Land cover transitions are the product of three underlying properties (Figure 1). First, the transition represents a discrete change of state. This means that if a location, such as a point, polygon, or pixel, at one time is in state A and and some other time in state B, then we can say that a state change has taken place. In some cases, land use and land cover transitions are rather more subjective in their classification than this statement implies. For example, a wooded forest lot may become the back yard of a luxury residence, and so change from forest to residential, changing use without changing cover. Nevertheless, we will assume that type transitions are discrete and definitive for the arguments in this work.

Secondly, land transitions have spatial location. For every transition between states, there is a geographic location that can be ascribed to that place. This information is not recorded when information is retained only about class changes, yet is important if land transition models are to simulate transitions, especially dynamically. In general, we can state that the places are not spatially independent. There is a strong degree of spatial autocorrelation between land transitions, some positive, and some negative. For example, in a rural area a small cross-roads may become the location for a cluster of residences. A single pixel may change from forest to urban. In the next step, land immediately adjacent to the buildings may be cleared of trees for gardens and farming. Thus the two transitions are directly spatially adjacent because they share a common origin in a sequence of events. At the broader scale, city edges are characterized by spatially autocorrelated transitions, as are resorts, farming at the periphery, and so forth.

Thirdly, land transitions are contextually linked. A simplistic model of land transitions, for example, could compute a transition matrix, select a state at random, draw a random number that matches a transition, select at random a pixel in the old state, and enforce change on that pixel. This ignores, however, the fact that the pixel selected for change has a spatial context, or a local "state neighborhood." Thus a pixel surrounded by eight adjacent pixels of the same state might be considered far less likely to change than a pixel surrounded by pixels all in different states.

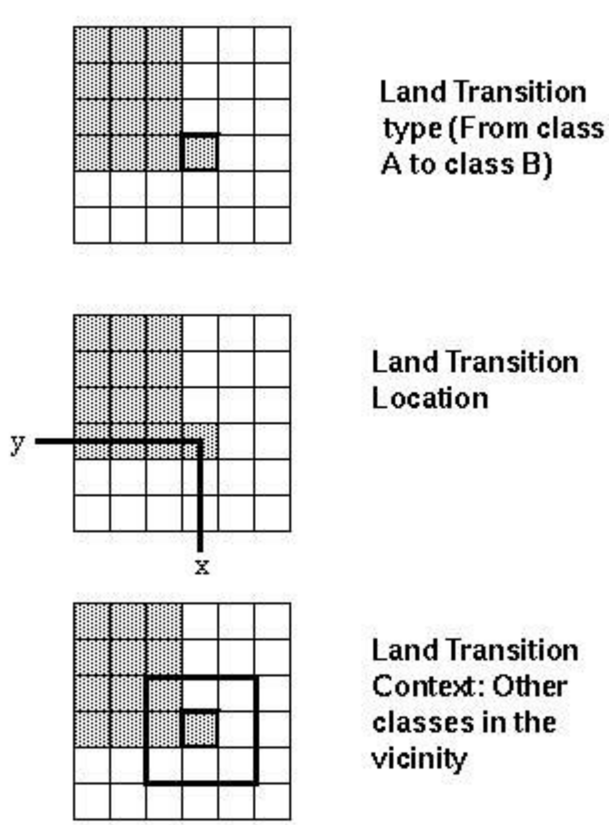


Figure 1: The three types of land transitions.

These three transition factors, what might be called *state transition*, *location*, and *spatial context*, should form the basis of a majority of transitions in a model of land transition, but obviously not all. Occasionally, particular transitions happen at random. Forest becomes wasteland because of a wildfire, or a pleasant spot in the forest is selected as the location for a residence. The transitions these changes initiate within the immediate vicinity are not random. The forest to wasteland example is unlikely to change to urban use without water, power or roads access, but may now be suitable for agricultural use as pasture. Thus once change has begun, however initated, a set sequence may be followed that is non random, both in further transitions on tha same location, and in adjacent transitions.

In addition, the motive force behind land transitions is usually associated with a single land cover type. In an ecological model of species, for example, an endangered or an invading species may be making all the spatial rules. In the broad land cover context, the dominant driving transition is the one converting unsettled land to human settlement, the urbanization land transformation. Prior work has used this approach to model urbanization from the past into the future. It is also possible to model a more full set of land use/cover transitions with similar methods, given that ultimately all land transitions are the result of the ongoing urbanization process itself.

Models of land transitions start by examining the state transition. A factor of interest here is that although for any two images at times 0 and 1 a state transition probability matrix ***T*** cab be computed, without further data it is impossible to state whether or not these transitions are stable. While it can fairly safely be assumed that the rows of transition probabilities sum to one, any given probability may actually vary randomly, systematically, or chaotically. Raising the matrix to a large power to produce a "static equilibrium state" therefore, is speculative to the highest degree. We propose that the probabilities are stable for short periods, but should be recomputed as many times as there are data available to do so.

Modeling Land Transitions

Land transitions fall into a simple modeling framework. Assuming ***n*** land cover/use classes at time zero and the same number of classes at time 1, then we can compute by direct measurement the transition matrix ***L***, which is an ***n*** x ***n*** matrix. For a set of locations, usually the locations of a regular array of pixels with matching extent and resolution in the two land use/cover images, for each element of ***L*** there are ***L[r ,c]*** transitions from type ***r*** to type ***c***. If we sum all counts of transitions along rows, then divide each entry by the sum, then what results is a matrix of transition "probabilities" ***T*** for the time step, that is we make the assumption that all land transitions are independently likely and that overall transition probabiltities are revealed over a large number of observations.

The ***T*** matrix can be generated by having at least two identical land cover maps at different time periods. For example, for the two symthetic images in figures 2 and 3, the matrix in figure 4 was generated.

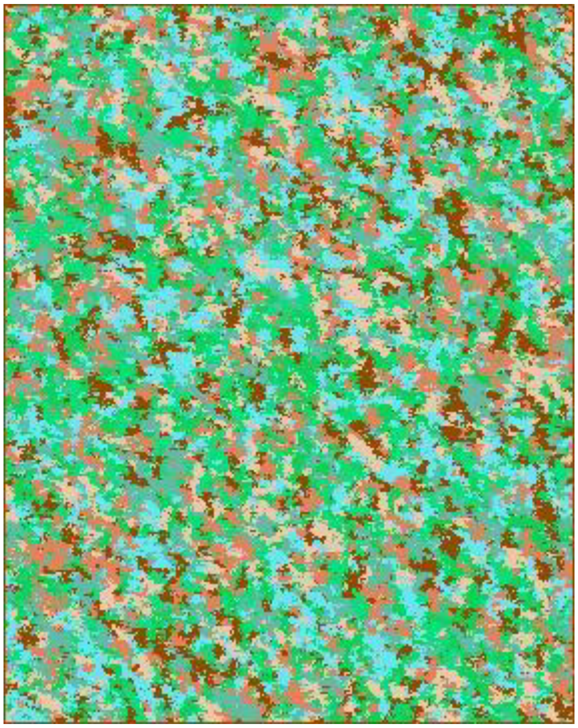


Figure 2: Synthetic Land Use at Time 0.

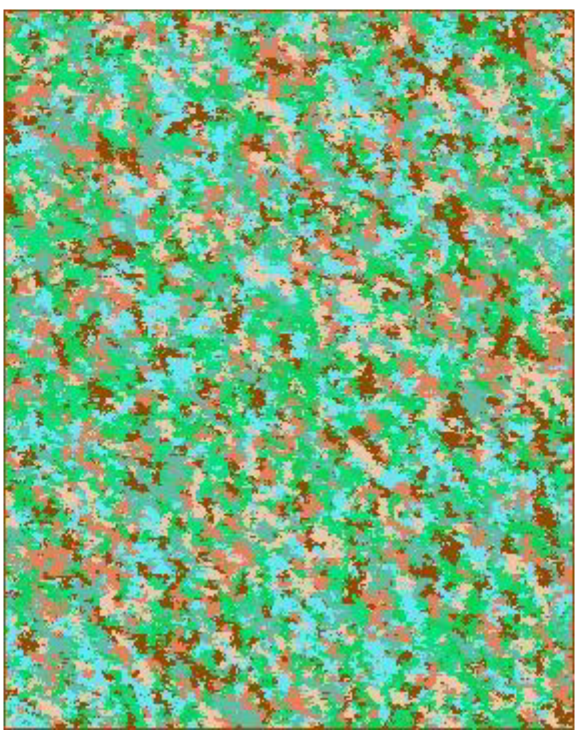


Figure 3: Synthetic Land Use at Time 1.

Figure 4: Land Cover Transition Matrix for figures 2 and 3. Counts and Probabilities

14786 [0.9936]	14 [0.0009]	23 [0.0015]	13 [0.0009]	17 [0.0011]	12 [0.0008]	16 [0.0011]
12 [0.0008]	14160 [0.9937]	14 [0.0010]	19 [0.0013]	12 [0.0008]	14 [0.0010]	19 [0.0013]
16 [0.0012]	5 [0.0004]	13696 [0.9943]	17 [0.0012]	11 [0.0008]	12 [0.0009]	18 [0.0013]
16 [0.0011]	13 [0.0009]	16 [0.0011]	14297 [0.9934]	17 [0.0012]	16 [0.0011]	17 [0.0012]
13 [0.0009]	14 [0.0010]	9 [0.0006]	9 [0.0006]	14016 [0.9954]	9 [0.0006]	11 [0.0008]
7 [0.0005]	11 [0.0008]	10 [0.0007]	7 [0.0005]	17 [0.0012]	14413 [0.9945]	27 [0.0019]
19 [0.0012]	18 [0.0012]	20 [0.0013]	14 [0.0009]	18 [0.0012]	18 [0.0012]	15336 [0.9931]

With only two land use/cover maps, obviously nothing can be determined about the statistical nature of the probabiltities. With three maps, a crude model of change or at least a variance estimate can be made.

The driving force of urbanization can be added by using an external model for these transformations, here we will use the Cellular Automaton Model, to generate a set N of newly grown pixels in a cellular landscape. The number of newly urbanized pixels is used as the driver for probabilities derived from the land transition matrix. For example, having normalized the probabilities in ***L*** to produce ***T***, then with x urban transition pixels, x = X/N calls can be made to the random number generator. The result is that further land transitions (excluding those related to urban) are generated within a single model iteration.

The Deltatron Approach

The approach adopted here is that of the deltatron. The deltatron is designed to enforce the following conditions:

1. That land transitions be considered to take place on a uniform spacing grid.
2. That transitions be between and among a finite set of states, where the number of states is small.
3. That the transition matrix accurately estimates land use state transition probabilities from observed counts.
4. That an external model be used to change the state of the dominant or driving class.
5. That there should exist considerable spatial autocorrelation in land transitions.
6. That these exists temporal correlation between land transitions.
7. That specific land transitions are influenced by context.
8. That land transitions happen to some degree at random, i.e. independent from the driving force.

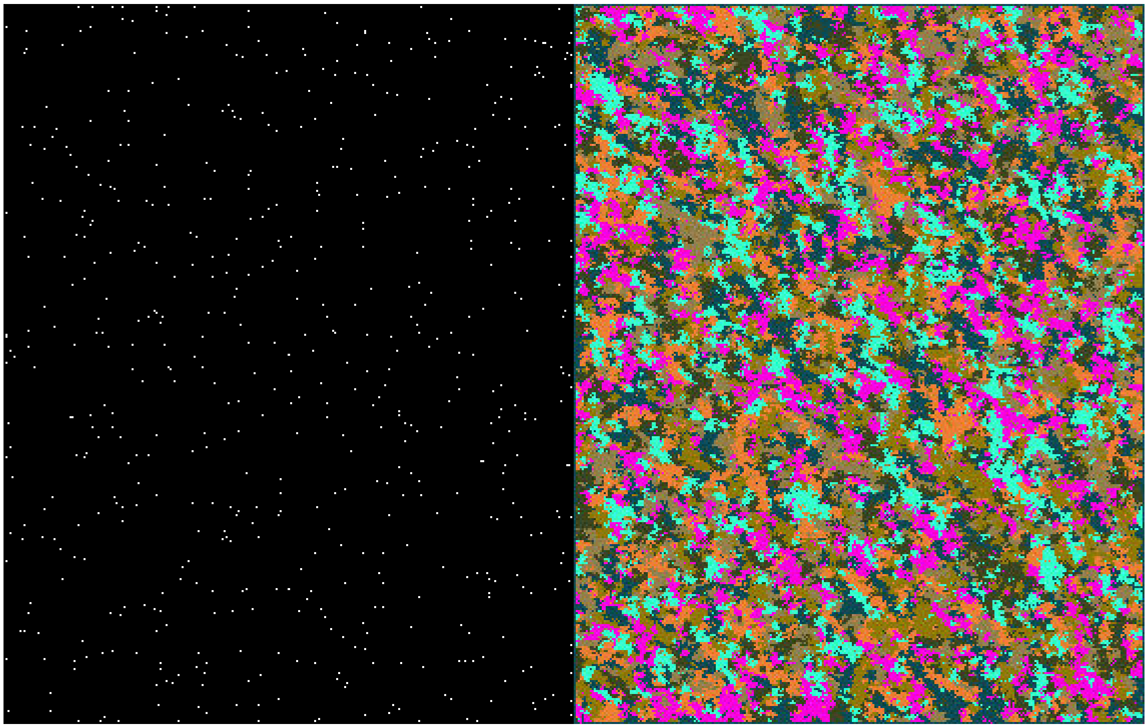
The deltatron is an artificial "agent" of change that has "life" in change space. Whenever a land cover transition takes place, a cell is "born." A deltatron attempts to change the land use/cover class of the immediate neighborhood. A deltatron will survive one iteration if in the next time step another transition of the same "to" type (i.e. same column in the transition matrix) happens in the eight-cell neighborhood of the transition. In this next step, no further transition can happen to the original pixel. Some deltarons are "killed" at random. Others die because no further change takes place in the vicinity. Thus eventually, even changed pixels can change at some time in the future.

The impact of the deltatrons is to act as cellular automata. Periods of growth and decline in change take place naturally. There are phase transitions in change, and the system seems to fluctuate between spatial structure (with fairly simple diffusion of growth outward from an initial point) and chaos. During periods of rapid driver impact (urban growth), obviously more deltatrons are created and they have more impact and interaction, resulting in a more spatio-temporally chaotic pattern. In periods of low growth, simple diffusion is normally follwed by death and the land use/cover change is "absorbed" by the landscape.

Synthetic Examples

So far, only synthetic examples have been programmed. We intend to use the synthetic examples to investigate the relationship between phase changes and driving factors. An effort has been made to keep the system closed. Change determinants are the driving number of newly urban pixels, the transition probabilities, and the number of permitted random deltatron births and deaths. These values will be determined in the implementation by using historical urbanization data for the San Francisco Bay data set. At leaset two historical land cover maps are available for this area. Work on a contemporary land cover classification will allow a third. Unfortunately, there are some inconsistencies in the land cover attribute classifications that must first be resolved.

Figure 4 (left) shows an animation of thirty Deltatron cycles for the synthetic land use data. The land cover maps undergoing the transitions are on the right. As is often the case in the real world, it is hard to determine that land cover changes are taking place at all from the spatial image. It was this fact that encouraged us to work in "delta" space to begin with.



Future Work

Our current project, Gigalopolis, has been extending and recalibrating the cellular automaton growth model in the Washington/Baltimore area. We will now move the model back to the San Francisco Bay area data set, to calibrate the land cover/deltatron extensions to the model. This will allow us to simulate present day land use from historical trends, and then to extend the model into the future to make further growth predictions. As before, we intend to animate the results.

In addition, we will be "scaling up" the entire model to cover the entire coterminous 48 United States at about 4km resolution. Sources of historical growth will be Atlases, USGS maps, the National Atlas, and other sources including the DMSP. Sources of land use in digital form will be the GIRAS files and the more contemporary land characterization images from the EROS Data Center. It is hoped to both assess the feasibility of scaling the model, and to make a set of projections for use in global change research.

References

- Clarke, K. C. "Project GIGALOPOLIS: Multiscale Calibration and Extension of a Predictive Land Characterization Model." (Proposal) USGS National
- Clarke, K.C., Hoppen, S. and Gaydos, L. (1997) "A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay Area". *Environment and Planning B: Planning and Design*, vol. 24, pp. 247-261.
- Clarke, K. C. and Gaydos, L. (1997) "Long Term Urban Growth Prediction Using A Cellular Automaton Model and GIS: Applications in San Francisco and

Washington/Baltimore. *International Journal of GIS*, Special Issue of Population Modeling and development, (Under Review).

- Clarke, K.C., Hoppen, S., Gaydos, L., (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.
- 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.
- Kramer, J. (1996) Integration of A GIS with a Local Scale Self-Modifying Urban Growth Model in Southeastern Orange County, New York." M.A. Thesis. Hunter College-CUNY.

MODELING URBAN LAND USE CHANGE:

The Next Generation of the California Urban Futures Model

John Landis and Ming Zhang

University of California, Berkeley

May 15, 1997

partial funding for this research was provided by the

University of California Transportation Center

The first generation of the California Urban Futures Model (Landis, 1994a, 1994b, 1995) achieved four significant advances in the field of metropolitan growth modeling. It was the first urban growth model which could be used to simulate how realistic regional and/or local development restrictions might impact future development patterns. It was the first operational urban growth model to be truly disaggregate--that is, not to rely on a zones. It was the first operational urban growth model to explore and model the crucially important role of private land developers in determining the future form of metropolitan growth. And, it was the first urban growth model to incorporate--indeed, depend on--GIS.

These advances notwithstanding, the original also suffered from some significant shortcomings. It was limited to residential development. Omitted from the model were methods for projecting and/or allocating future industrial, commercial, and public activities. A second limitation, which was a direct outgrowth of the first, was that the model did not allow different activities to bid against each other for appropriate sites. Unless explicitly prohibited, the "best" sites (that is, those sites which were the most profitable to develop) were always reserved for residential development. Third, and perhaps most critically, the rules used to allocate future development had never been *calibrated* against historical experience. Instead, residential growth was allocated to sites based on the difference between observed housing prices, and "best-practice" housing development cost functions. Finally, real estate prices, to the extent that they played a role in the CUF Model, were entirely exogenous. New residential development that could not be allocated simply *spilled-over* into other jurisdictions (subject to user-specified limits on spillover), rather than feeding back into the allocation process through higher prices.

The second generation of the California Urban Futures Model (CUF II) tries to remedy these shortcomings. Unlike the first generation. it:

* Includes multiple sectors;

* Allows different land uses to bid against each other for preferred sites;

* Is calibrated against recent experience; and,

* Incorporates a "pseudo-pricing dynamic" into the development spillover process.

This article explores the development and calibration of the CUF II Model. It begins with a look at recent developments in large-scale urban modeling. Next, it reviews the overall logic of the CUF II Model. Third, it explains the development and calibration of the Land Use Change Model, the newest component of the CUF II Model. It concludes by reviewing the CUF II's advances as well as its remaining deficiencies. A companion article, describes the use of the CUF II Model for simulating county and regional development policies.

RECENT DEVELOPMENTS IN METROPOLITAN ACTIVITY MODELING

As two recent reviews by Wegener (1994, 1995) point out, the state-of-the-art of operational metropolitan activity models is evolving at moderate pace. Two developments of recent years--one theoretical, the other data-oriented--are combining to help metropolitan growth modelers produce ever-more explicit models of urban activity patterns. The first of these is the incorporation of random-utility theory to predict generalized location and travel choices based on observed household, firm, or traveler behaviors. The ability to model metropolitan-scale activities as combinations of individual location and travel choices represents a significant step forward in the theoretical development of urban models.

A second advance is based on the growing availability of longitudinal, micro-scale data, and on the growing power of GIS-based software to analyze such data. Activity data tied to specific locations in space (either through geo-referencing or address-matching) is now available in many locations. This makes it possible to represent the metropolitan area as a collection of individual sites or locations, rather than as zones. It also makes it possible to track and analyze actual patterns of metropolitan land use, demographic, or economic change, instead of relying on changes in zonal aggregates. Building on such newly available data sources first Batty (1992, 1995), and more recently Clarke (1996), have begun to insert processes of spatial change (such as cellular growth and diffusion) into urban activity models.

Despite these advances, a number of significant problems remain. The first is the absence of actual--which is to say, observed-- land or building prices. Site prices are typically determined endogenously in most urban growth models, usually as travel-based shadow prices, or else indirectly through the process of market-clearing. This limits the ability of most metropolitan activity models to simulate the effects of realistic land use policies and regulations, or of non-transportation infrastructure investments. Until urban activity models can be calibrated against the characteristics of observed land use transactions--including location as well as price--their usefulness for real world applications will remain suspect.

A second problem area concerns the modeling of economic activity, particularly jobs. Most existing metropolitan models function adhere to some form of export-base framework. The total demand for regionally exported goods and services is determined exogenously as a model input. Once determined (and converted to jobs), regional economic activity is then sectorally disaggregated into primary and secondary industries, as well as spatially disaggregated by zone. The processes by which actual job changes really occur--through the growth or shrinkage of existing industries, through the birth of new industries, or the decline of older ones, is ignored. Similarly ignored are the non-transportation cost reasons why industries locate where they do.

A third problem area is the continuing (and virtually complete) lack of supply-side information in most urban models. Except for those areas that are absolutely precluded from development (by virtue, for example, of being underwater), the suitability of the built and natural landscape to support new or additional development; or, the capability of the landscape to support development at differing intensities, is completely missing from most all metropolitan activity models. The characteristics of the existing landscape, be it natural and undeveloped, or already developed, are widely presumed to be irrelevant to the amount or location of future activities or land uses. This limitation has meant that urban models can not be reliably used for project-based environmental, social or fiscal impact analysis, or for cumulative impact assessment.

A related problem, first identified by Lee (1973), is the insensitivity of most urban models to many types of policies. Today's generation of urban models are useful for analyzing the travel, congestion, air quality, and perhaps, regional economic implications of new highways and/or transit extensions, but not for much else. Because they typically exclude site-level environmental or regulatory information, today's metropolitan activity models remain incapable of analyzing or simulating local land use and environmental regulations. Indeed, most metropolitan growth models are generally blind to local government boundaries, issues, or policies. The ability to model how local policy changes will affect activity patterns within particular municipal boundaries, let alone model the inter-municipal spill-over effects of policies, is largely missing from today's urban models.

THE LOGIC OF THE CUF II MODEL

The original CUF model was developed to fill some of these gaps. As noted by Wegener and others, it sacrificed comprehensiveness and theoretical elegance in favor of spatial detail and an ability to simulate local policy initiatives. The newest version of the model maintains the same policy-focus and level of spatial detail as the original, while plugging some of its theoretical holes.

Despite its more developed theoretical base, the CUF II Model is conceptually simpler than its predecessor. As shown in Figure 1, the new version consists of three, not four components:

1.Activity Projection: The first component consists of a series of econometric models used to project future households and employment by jurisdiction at ten year intervals. Households are divided into owners and renters. Employment is separated into ten sectors. The development and calibration of the different employment models is detailed in Landis, et.al., 1997.

2.Spatial Database: The second component of the CUF II model consists of a GIS-based database of Developable Land Units (DLUs). As in the prior version, DLUs are potentially developable*or redevelopable*sites. In the original CUF model, DLUs were generated as the spatial union of multiple GIS layers. The resulting DLU database consisted of thousands of unique polygons, each defined as different along some attribute dimension than its immediate neighbors.

In the current version, DLUs consist of one-hectare (100m x 100m) grid cells, which may or may not be uniquely different from adjacent grid cells. The shift from polygons to grid-cells was made for a number of reasons. First, the analytic power of grid-based GIS procedures has advanced considerably since the first version of the CUF model. Second, data on existing and historical land uses are collected and tabulated at the hectare level by the Association of Bay Area Governments (ABAG). This data is essential to the development and Calibration of the Land Use Change Model, described below. Finally, the hectare grid cell has about the right level of resolution for analyzing pattens of development and land use change. It is sufficiently fine-grained to observe small-scale land use changes, but not too fine-grained so as to obscure those changes with "noise."

The Spatial Database includes ten data layers: i) 1985 and 1995 land uses, by hectare, as identified by the Association of Bay Area Governments; (ii) Percent slope, as estimated from

USGS Digital Elevation Model (DEM) data; (iii) Publicly-owned or controlled lands; (iv) Wetlands, as identified from the National Wetlands Inventory; (v) 1990 City boundaries, obtained from 1992 TIGER files; (vi) 1990 Spheres-of-influence, as digitized from paper maps provided by County LAFCOs; (vii) Urbanization and agricultural land quality, as determined from digital maps provided by the California Farmland Mapping Project; (viii) 1990 General plan designations, as digitized from paper maps provided by Contra Costa, Sonoma, and Solano Counties; (ix) Major highway rights-of-way and interchange locations, as obtained from 1992 TIGER files; and (x) Major rail transit rights-of-way and stations as digitized from paper maps. As noted in Chapter Three, several additional data items are developed from these ten data layers.

3.The Land Use Change Model: The Land Use Change Model is the heart of the CUF II Model. It is a series of equations that relate hectare-scale land use changes (between 1985 and 1995) to more than two-dozen site and community characteristics, including: local population and employment growth; proximity to regional job centers; site slope; whether the site is within or beyond city boundaries or spheres of influence; the uses of surrounding sites; the availability of vacant land; site proximity to freeway interchanges and transit stations; and site proximity to major commercial, industrial, and public land uses.

Because land use change is a discrete rather than continuous phenomenon, the various equations are estimated using a multi-nomial logit procedure. Nine types of site-level land use changes are considered: (i) undeveloped to single-family residential use; (ii) undeveloped to apartment use; (iii) undeveloped to office or retail use; (iv) undeveloped to industrial use; (v) redevelopment from another developed land use to residential use; (vii) redevelopment to commercial use; apartment use; (vii) redevelopment to industrial use; (xiii) remain*undeveloped*; and (ix) remain in initial*developed*use. Separate logit models are calibrated for new development and for redevelopment, and for each Bay Area county.

The resulting model parameters can be combined to calculate land use transition probabilities (e.g., the probability that a specific vacant site will be developed in residential use). These probabilities, in turn, may be interpreted as "bids" for development or redevelopment. Projected new development can then be allocated to sites according in order of their bid scores. Bid scores vary by site as well as by potential use. This means that different uses (e.g., single-family residential, apartments, commercial and industrial uses) can effectively bid against each other for specific sites. In this way, the Land Use Change Model incorporates competition between uses and between sites.

This "bidding" framework allows for two types of spillover. If, for example, there are insufficient sites to accommodate projected residential development, but sufficient sites for commercial development, residential growth may be allocated (or spillover) to sites whose "highest-and best-use" (as represented by their bid scores) would otherwise be for commercial development.

Bid scores can also be evaluated across jurisdictions. This facilitates inter-jurisdictional spillover. Activities for which sites are not available in one jurisdiction can spillover into other jurisdictions. This is the most common form of spillover. Alternately, activities may be allocated to sites regardless of jurisdictional boundaries.

Like its predecessor, the CUF II model can be used to simulate future policy scenarios. Unlike the original CUF model, however, which was limited to testing regulatory scenarios, the CUF II Model can also be used to simulate investment scenarios, such as the construction of new freeways or transit systems. The results of these simulations show the locations and patterns of development, in addition to the total amount and density.

MODELING URBAN LAND USE CONVERSION: A DISCRETE CHANGE APPROACH Viewed from an airplane or a satellite, processes and forms of metropolitan land use change often appear to follow regular and explainable patterns. Lower-density residential development extends outward as farm or resource lands at the urban edge are gradually developed. Industrial areas develop around major inter-urban transhipment facilities such as seaports, airports, railroads, or highways. Commercial and office development occurs around key intra-metropolitan transportation nodes. Apartment complexes are developed near central city and suburban central business districts.

What appears be regular and explainable--perhaps even inevitable--at the metropolitan scale, is not quite so regular when viewed at the level of the individual parcel or site. There are many reasons why and when individual sites change land use. Some reasons are well-known and consistent with established theories of metropolitan growth: A new highway makes a low-priced agricultural site accessible to regional job centers. New residential development exceeds the market size threshold required to make a nearby retail center economically viable. Restrictive land use controls enacted by higher-income municipalities force new development to be displaced outward.

Other reasons are more idiosyncratic. A farmer who prefers to continue farming refuses to sell his land to a subdivider despite being offered a terrific price. Excessive exactions or uncompromising neighborhood opposition prompt a developer to skip over a preferred site in favor of a one which is just easier to develop. A retail developer misjudges the market and builds a shopping center before there is sufficient residential demand. One land speculator in need of ready cash sells to another who has deeper pockets. A pro-growth majority planning commission gives way to a slow-growth majority.

Over the long-run, perhaps 20 or 30 years, many of these idiosyncratic reasons give way to the logic of the market. The heirs of the reluctant farmer finally decide to sell to a developer. Rising land and property values allow developers to pay for required infrastructure or environmental mitigation, negating previous fiscal or environmental constraints. Commercial areas establish themselves, generating real agglomeration economies. Looked at from above, or better yet, in retrospect, the pattern of metropolitan growth seems inevitable.

Far from being inevitable, the pattern of metropolitan growth is actually*path-dependent*. Which particular sites were developed (or not developed) in the 1960s determine which sites will be available for development or redevelopment in the 1980s. If the truculent farmer*had*sold his land in the 1970s, the metropolitan area might have looked different in the 1990s.

Because metropolitan growth actually occurs as a cumulative, path-dependent process of individual parcel changes, it has proven to be extremely difficult to model statistically. Conventional--which is to say, regression-based--statistical approaches are not particularly appropriate to the task of modeling site-level land use change, or for incorporating site-level spatial effects such as proximity. Instead, urban modelers have chosen to focus on aggregate patterns of urban growth, usually by studying land use change (or, more specifically changes in level of activities which use land) at the zonal level. The characteristics of individual sties, as well as the factors which shape the motivations of the buyers and sellers of those sites, then conveniently disappear.

The Logit Framework

Choices or changes between categories, for example, the change that occurs when a vacant parcel is developed to a residential use, are more appropriately modeled using a non-linear logistic, or*logit*model. Logit models typically come in two forms:*binomial*, meaning there are only two choice or change possibilities; and*multi-nomial*, meaning that there are three or more choice or change possibilities. Multi-nomial logit models are typically much harder to calibrate and interpret than binomial models. Because we are modeling changes into multiple land use categories (e.g., single-family residential vs. multi-family-residential vs. commercial, vs. industrial.), all of the models presented below are multi-nomial in form.

They are also *non-ordinal*. This means that the categories which make up the dependent variables can not be ordered or ranked. As applied to processes of land use change, this means that no one type of land use change can be presumed, a priori, to be universally superior to another. The extent to which commercial uses, for example, are judged superior to residential uses, is determined exclusively by the data and the resulting model coefficient estimates. The fact that the models are non-ordinal means that various land uses can effectively "bid" against each other on a site-by-site basis.

Regardless of whether it is ordinal or non-ordinal, the multi-nomial logit model takes the following general form:

$$Prob[i|l] = \left\{ \frac{\exp(\beta_0^l + \beta_1^l x_{i1} + \beta_2^l x_{i2} + \ldots\ldots + \beta_m^l x_{im})}{\Sigma_{l=1,L} \exp(\beta_0^l + \beta_1^l x_{i1} + \beta_2^l x_{i2} + \ldots\ldots + \beta_m^l x_{im})} \right\}$$

where: Prob[i|l] is the probability that each grid-cell site *i* is developed or redeveloped to use *l*;

x_{im} are explanatory, or independent variables for each site *i*;

β_m are the logit coefficients (to be estimated) associated with land use *l* and variable x_m;

and L is the full set of land use changes.

Multi-nomial logit models have primarily been used to analyze the behavior of individual decision units at a single point in time. As explained by McFadden and Domenich (1975), the use of a the logit estimator to model individual choices is based on three assumptions. The first is that when faced with complete information, individuals will rationally choose the alternative that maximizes their own utility. A second assumption is that the commodity space which describes the alternative set must be differentiable and convex (Fisher and Nijkamp 1985). A third assumption--one which is usually treated as implicit--is that individual choices are independent. What this means is that the decisions made by one individual *neither do nor can* affect the decisions made by other individuals. This last assumption rules out the possibility of strategic behavior.

Adapting the multi-nomial logit model to the task of explaining land use *change* requires some additional assumptions. We begin by assuming that the decision to develop a previously vacant site (or to redevelop a previously developed site) will be based on a rational evaluation of the prospective profit or rent associated with different potential forms of development.. For a given site (denoted by the subscript *i*), let *R*(*y*/*i*) denote the profit potential associated with a particular land use change, *y*. An alternative land use change, *y** will occur if and only if :

$$R(y^*/i) > R(y'/i) \text{ for all } y^* y'.$$

That is, the site will be developed into the land use that generates the highest potential profit. The profit potential associated with each choice, $R(y/i)$, is determined by a set of site attributes. Some attributes are observable. Others are indeterminate or unobservable. Because some attributes are unobserved, the land use change function is assumed to be probabilistic. Let $Pr(y/i)$ denote the probability that choice y is made for site i. Under assumptions of profit-maximization:

$$Pr(y/i) = \text{Prob}\ [R(y/i) > R(y'/i) \text{ for all } y' = y]$$

Under the simplest assumption that the unobserved attributes are independent and identically distributed according to a Gumbel Type I extreme value distribution, this probability takes the form of a multi-nomial logit function. (See McFadden 1974, for the mathematical proof of this derivation). Much as random utility theory underlies the use of the logit function to model consumer choice, we use the idea of a "random" profit function to develop models of land use change.

Transforming consumer-based random utility theory into developer-based random profit theory requires overcoming two problems. The first is theoretical. It involves the assumption that land developers act independently of each other--that is, that each developer or landowner *independently* appraises the profit potential for every site, and bids accordingly. As noted above, this assumption rules out the possibility of oligopolistic (whereas groups of landowners or developers act in concert) or strategic behavior (whereas one developer acts primarily to pre-empt or manipulate another). The problem is that landowners and developers *do* engage in oligopolistic and strategic behavior. And so, for that matter, do many land sellers.

Perhaps the more appropriate question is not whether land sellers, landowners, and developers engage in strategic behavior (we assume they do); it is whether that behavior is likely to succeed. To the extent that land development has been shown to be no more profitable over the long-run than other businesses, the answer to this second question is probably no. Competition, we assume, levels the playing field and makes the expected return (or profitability) associated with strategic behavior close to zero.

A second problem stems from the nature of the observations used to calibrate the models, below. It revolves around the question of *agents*. In the case of conventional discrete choice analysis, the agent is the individual or household. Consider the case of commuters comparing alternative work trip modes, or of households trying to decide where to live. In the commuter case, each traveler faces a series of mode choices (e.g., driving, walking, or taking the bus), all of which can be decomposed into a comparable set of attributes (e.g., travel time, wait time, travel cost). In the household location case, each household faces a series of residential choices, all of which can also be decomposed into comparable attributes (house size, neighborhood, distance to work, school quality, etc.). Each traveler chooses the work mode, which, based on its attributes, maximizes his or her utility. Similarly, each household chooses a house and location, which, based on their joint attributes, maximizes its utility. In both examples, an identifiable agent confronts and makes real choices.

Now consider the case of site level land use change. The agent in this case*should be*the site owners (or developers with site control). Each owner is confronted with the decision of whether or not to initiate a land use change. The factors influencing that decision, will include, among others, the attributes of the site. Following the logic identified above, each owner should make the land use change decision (including the possibility of no change) which maximizes their profits.

Yet as we note below, the unit of analysis (or observation) in this research is the site, not the developer or landowner. And while we have reasonably complete information on the characteristics of sites, we lack information regarding the characteristics or motivations of land owners and developers. Put another way, we lack agents. To overcome this problem, we again invoke the idea of competition. We argue first, that given a highly competitive market and few barriers to entry, the agent doesn't matter. Whether a particular developer is well-capitalized or poorly-capitalized, whether they specialize in residential development or retail development, whether their experience is local or national; in a competitive market, these factors are likely to be of far, far less importance than the demand for urban development and the availability of appropriate sites.

Unit of Analysis

The parcel is the near-ideal unit of analysis for studying land use change. Parcels have area, location, single uses, and best of all, are the basis of all land transactions. Regrettably, complete (digital) parcel maps are not yet available for any Bay Area county for any year.

What*is*available, is a region-wide database of dominant land uses by organized by hectare. As compiled by the Association of Bay Area Governments (ABAG), this database is organized according to the Anderson (197-) land cover classification system. To make the modeling process more manageable, we collapsed the 100-plus land use categories contained in the ABAG land use database into seven: (i) undeveloped; (ii) single-family residential; (iii) multi-family residential; (iv) commercial; (v) industrial; (vi) transportation; and (vii) public.

Hectare grid-cells have both advantages and disadvantages as units of analysis. They are small enough to capture the detailed fabric of urban land used but large enough to avoid problems of data "noise." And, since they are fixed, changes and trends across time can be easily identified. On the negative side, they lack physical or legal reality. Unlike parcels, they are not transacted. Nor are they directly regulated. Thus, they are not themselves the subject of development or redevelopment decisions.

The nine-county database includes nearly 1.8 million grid-cells. Even when stratified into smaller county subsets, the database is too large to analyze within a multi-nomial logit framework. To make the analysis more manageable, we eliminated those grid-cells which we believed*were extremely unlikely*to have changed land use between 1985 and 1995. These included grid-cells which were known to be wetlands, grid-cells more than 50 kilometers from a highway, and grid cells more than 30 kilometers from a highway with a slope exceeding 20%. The effect of this screening was to reduce the quantity of land use changes to be modeled to a computationally manageable level.

MODEL SPECIFICATION AND CALIBRATION

Two sets of logit models of land use change are calibrated below. The first examines the determinants of land use change among undeveloped sites between 1985 and 1995. The second looks at the determinants of land use change among previously developed sites. Both types of models follow the same general form:

Pr[Land use change$_{ijkl}$] = f { initial site use$_i$, site characteristics$_i$, site accessibility$_i$,

community characteristics$_j$, policy factors$_{ij}$,

relationships to neighboring sites$_i$}

where: Pr[Land use change$_{ijkl}$] indicates the probability that site *i* in community *j* changed from land use *k* to land use *l* between 1985 and 1995;

i ndicates each hectare grid cell;

k, the initial (1985) land use of site *i*, is either undeveloped or developed;

And *l*, the terminal (1995) land use of site *i* is either single-family residential, multi-family residential, commercial, industrial, or else is unchanged from the initial use.

The two sets of models were calibrated for every Bay Area county except San Francisco. As noted above, the calibration datasets includes all developed sites in the Bay Area as of 1985, and a 50% superset of 1985 vacant sites most likely to be developed. Table 1 presents a frequency of the calibration datasets organized by county and by initial and terminal land use.

Table 1

The Dependent Variable: Categories and Levels of Land Use Change

Most of the sites that were vacant in 1985 were still vacant in 1995. The percentage of undeveloped sites in 1985 which were also undeveloped in 1995 ranged from a low of 94.3% in Contra Costa County, to a high of 98.9% in Marin County. Of the initially vacant sites that *did* change land use between 1985 and 1995, the largest share changed to single-family housing. The share of 1985 vacant sites converted to single-family use by 1995 ranged from a low of .9% in Marin County, to a high of 5.2% in Contra Costa County. Commercial and industrial development accounted for much less vacant land conversion than did new residential development. The share of 1985 vacant sites converted to commercial uses by 1995 ranged from a low of .17% in Napa County, to a high of 1.2% in Alameda County. The share of 1985 vacant sites converted to industrial uses also varied widely by county--from a low of .04% in San Mateo County, to a high of .54% in Alameda County. New apartment construction throughout the Bay Area was relatively meager between 1985 and 1995. The share of 1985 vacant sites converted into apartment use ranged from a low of .01% in Napa and Solano counties, to a high of .1% in Alameda County.

Redevelopment rates varied even more widely among counties than rates of new development. Redevelopment rates --that is, the area of sites which were redeveloped between 1985 and 1995 as a share of all developed sites in 1985--varied from a high of 5.2% in Alameda County, to a low of less than .1% in Marin County. As might be expected, redevelopment rates between 1985 and 1995 were higher in the Bay Area's urbanized counties (e.g., Alameda, Contra Costa, San Mateo, and Santa Clara), and lower in the its rural and suburban counties (Napa, Solano, and Sonoma). Marin County, was exception to this. Despite the county's older, more established status, only 14 hectares (about 35 acres) of redevelopment occurred in Marin County between 1985 and 1995.

Most redevelopment activity, regardless of county, took the form of residential redevelopment; that is, redevelopment from some other urban use, to housing. The share of developed sites in 1985 which were redeveloped to residential use by 1995 varied from a high of 4.8% in Alameda County, to a low of .1% in Marin County. Measured in absolute terms instead of percentages, the greatest amounts of residential redevelopment occurred in Santa Clara (2,340 hectares) and Alameda (1,096) counties. Santa Clara County also led the Bay Area in hectares of commercial and industrial redevelopment. Five hundred hectares (about 1,250 acres) of urban uses in Santa County were redeveloped to office, retail, and business park uses between 1985 and 1995. Another 71 hectares were redeveloped to industrial uses. Commercial redevelopment activity between 1985 and 1995 also exceeded 100 hectares in Alameda, Contra Costa, and Sonoma Counties. Other than Santa Clara, Sonoma County was the only Bay Area county in which there were significant amounts of industrial redevelopment 1985 and 1995.

Independent Variables: The Many Determinants of Land Use Change*:*

The general logit model specified above includes six sets of independent variables. They are:

1. The initial site use.

2. The demand for particular types of land uses in the general area.

3. The generalized accessibility of the site to other activities.

4. The nature and extent of any physical constraints, and/or the cost of overcoming those constraints.

5. The nature and extent of local policy constraints (such as zoning controls) which limit or restrict development, and/or the difficult of changing those constraints.

6. The existence of any positive and/or negative influences exerted by neighboring sites.

Not all of these factors can be conveniently or comprehensively measured. Site-level demand is particularly difficult to measure, as are certain types of policy constraints. Current and reliable data describing specific factors such as zoning, impact fees, traffic congestion, and infrastructure availability are hard to acquire and expensive to encode in a form convenient for use with statistical models. In a region with as many people (almost six million) and local governments (over 100) as the Bay Area, data collection and common encoding is particularly difficult.

The spatial scale of analysis also differs for different factors. Population and job demand are commonly measured and projected at county or jurisdiction level. Accessibility, by contrast, is commonly measured at the level of the traffic analysis zone (TAZ). Neighborhood effects, development constraints, and externality effects are more disaggregate still; if they are measured, it is usually at the site or parcel level. Modern relational databases and GIS packages enable analysts to overcome some but not all of these variations in spatial scale.

The following sections describe each of the independent variable sets:

1.Initial Site Use: The models that follow differentiate between land use change that occurs to previously vacant sites, and land use change that occurs to previously developed sites. The initial land use in the former set of models is all the same. In the latter set of models, sites may be initially developed in residential use (single-family or multi-family), commercial use (retail or office), or industrial use.

Conventional urban economics holds that commercial and industrial uses are generally of a "higher order" than residential uses. That is, they are capable of generating higher land rents. To the extent that this generalization holds true, previously-developed residential sites should, all else being equal, be more likely to be redeveloped into commercial or industrial uses. Conversely, previously-developed commercial or industrial sites should be less likely to be redeveloped into residential use.

2.Demand Factors: The probability that a vacant sites will be developed, or that previously-developed sites will be redeveloped, should depend in large measure on the strength of the demand for space. All else being equal, we would expect land use change to be more frequent in growing cities, and less frequent in declining cities. We measured demand in two ways: (i) as the rate of household growth or change during the previous five years (1980-85); and, (ii) as the rate of job growth or job change during the previous five years. Both variables were measured at the city level; all sites within a city were presumed to be subject to comparable demand pressures. Assuming that population and employment growth causes land conversion (and not vice-versa), we expected the estimate coefficients of both variables to be positive.

Too other general demand measures were also included: (iii) the initial number of households in the city (as of 1985); and (iv) the initial number of jobs in the city, also as of 1985. For reasons that are not exactly clear, population growth in the Bay Area during the 1980s favored newer smaller cities over older, larger cities. All else being equal, we would thus expect sites in larger cities to be less likely to either change land use or be developed than sites in smaller cities. This suggests that the estimated coefficient associated with the number of households in each city should be negative.

The size-effect of a city's employment base may be either positive or negative. On the positive-effect side, there may be agglomeration economies associated with larger employment centers. This would tend to make nearby undeveloped sites more attractive, thereby increasing their probability of being development. Moreover, recent employment growth in the Bay Area, unlike recent population growth, has been focused in a few large job centers. This would also suggests that the relationship between the size of a particular city's employment base, and the probability of a site within that city being developed or changing use is likely to be a positive.

On the negative effect side, land and commercial space prices are likely to be higher in cities with larger economies than in cities with smaller economies. To the extent that employers are drawn to less expensive land, the effect of employment size on land use change may be negative.

We included one final variable in the various models to test the hypothesis that the relationship between the number of jobs and households in a community somehow affects the likelihood of land use change. Depending on the terminal use, theory indicates that this variable could cut both ways: To the extent that cities prefer to attract a balance of jobs and housing, higher jobs-housing ratios might be negatively affect the likelihood that a site be developed in residential use. Conversely, because of agglomeration economies, new jobs might be attracted to already job-rich communities, thereby boosting the likelihood that a particular site be developed in commercial use.

3.Accessibility and Distance Effects: Starting with von Thunen, urban economists have argued that the demand for sites (as measured by land prices and densities) should be greatest near major city centers, primarily for reasons of minimized worktrip transportation costs. The San Francisco Bay Area has three regional employment centers (San Francisco, Oakland, and San Jose) and many more subcenters (e.g., Walnut Creek, San Ramon, Pleasanton, Fremont, Santa Clara/Sunnyvale, Palo Alto, San Mateo, Hayward, Berkeley, Richmond, Fairfield, Vallejo, and Santa Rosa). To capture any potential regional accessibility affect, we used GIS to measure the euclidean distance from every developable and redevelopable site to downtown San Francisco and downtown San Jose. To the extent development really does favor*closer-in*locations, we would expect the estimated coefficients of these two measures to be consistently negative.

Accessibility can be also be measured more generally. Regardless of trip destination or purpose, activities located near major freeway interchanges or transit stations have a higher level of generalized accessibility than activities located farther away. Because of this, we would expect such sites to be in greater demand, and thus, to face greater development and redevelopment pressures. To test this hypothesis, we measured the aerial distance from every site to the closest freeway interchange and/or BART stations. To the extent that proximity to regional transportation facilities encourages land use change, we would expect the coefficients of these two measures to be negative.

4.Physical and Cost Constraints: The physical characteristics of a site may present absolute or relative constraints to its development. Sites which include permanent wetlands are absolutely constrained from development. Sloped sites face relative constraints: they can be developed or redeveloped but typically at a higher cost than flat sites. To develop sites far from existing urban services (e.g., roads, sewer and water service, and electrical and telephone service) requires either that those services be newly provided, or that they be extend from existing service areas. Either way, the necessity of providing services substantially raises the increases the cost of developing vacant land at the urban fringe.

Sites identified as permanent wetlands under the National Wetlands Inventory are excluded entirely from this analysis.

Site slope is identified through the use of five dummy variables, corresponding to five slope classifications: (i) 0-2% slope; (ii) 3-5% slope; (iii) 6-9% slope; (iv) 10-15% slope; and (v) 16% or greater slope. Site slopes were originally estimated from U.S.G.S. DEM (Digital Elevation Model) data files. Because of the higher costs associated with hillside development, we generally expect to observe a negative relationship between the dummy variables denoting steeper slopes and the probability of site development. The dummy variable indicating a 0% slope was purposely omitted in order to guarantee a unique solution.

Developers who propose projects in existing urban areas (sometimes called "infill") are sometimes able to take advantage of existing road and transit capacity as well as hook-up to existing water and sewer lines, and to electrical and telecommunications grids. All else being equal, this reduces total development costs. Developers who propose projects at the urban edge may also be able to take advantage of existing infrastructure capacity; although in practice, significant and costly additions are often required as a condition of development approval. Developers who pioneer entirely new areas, or propose projects beyond the urban fringe either have to provide their own infrastructure capacity, or else pay to extend existing services to their projects. Either way, the costs of providing services to far-flung projects tend to be very high.

To try to capture this effect, we used GIS to measure the linear distance (in 200 meter increments) from each site to the nearest sphere-of-influence boundary. The greater this distance, we assume, the higher the cost of providing required infrastructure and urban services. Thus, we would expect to observe a negative relationship between the measured distance between a particular site and the closest sphere-of-influence boundary, and the probability that the site is developed.

The costs of providing infrastructure and essential urban services varies by use and jurisdiction as well as with distance. Some jurisdiction impose more costly and extensive infrastructure standards than others. Similarly, some jurisdictions impose more onerous standards on certain types of development. Finally, most Bay Area jurisdiction assess impact fees on new development. Because state law governing the setting of impact fees requires only that there be a "rational nexus" between the fee amount and the impact, fee amounts can vary widely between jurisdictions, between different uses, and even according to the location within a particular jurisdictions.*Because we were unable to assemble a complete and reliable schedule of impact fees for different uses in all Bay Area communities, we did not include impact fees (or exactions) in the model.*

5.Policy Constraints: Most constraints to development are political, not physical. Just about every local governments in California utilizes zoning to stipulate which uses and densities are permitted where. Similarly, California municipalities are supposed to designate their ultimate "build-out" boundaries as "spheres-of-influence." New development is supposed to be channeled to sites within sphere-of-influence boundaries, and steered away from unincorporated areas outside sphere boundaries. And increasing number of California jurisdictions are moving to limit the development of farmlands, primarily as a means of preserving open space.

Policy constraints are rarely absolute. Zoning can be, and is frequently changed. Indeed, experienced developers typically look for under-zoned properties in the hope of changing their zoning to a higher, and thus more profitable use. Likewise, sphere-of-influence boundaries can be easily extended. And incentive programs such as Williamson Act contracts intended to protect farmland have so far failed to catch on.

Because of the cost and difficulty of obtaining, digitizing, and coding accurate, up-to-the minute zoning maps for every jurisdiction,*we did not include zoning categories in the model*. We did, however, determine whether each site was within a current sphere-of-influence boundary. All else being equal, and despite the laxness with which current sphere-of-influence requirements are being implemented, we expect that vacant sites outside sphere boundaries would be less likely to be developed and thus change use.

Using digital maps of farmland quality published by the California Farmland Mapping Project, we determined whether undeveloped sites were located on land classified as: (i) prime farmland; (ii) farmland of state or local importance; (iii) farmland being cultivated with a unique crop; and (iv) livestock grazing land. Only the*prime farmland*classification was included in the model as a variable.

6a.Adjacent Use Effects: All else being equal, we would expect site land uses to be strongly affected by the pattern of neighboring or adjacent uses. We would expect, for example, that a vacant site surrounded by residential uses would be more likely to be developed into residential or retail use than into office or industrial use. Likewise, we would expect that a vacant site surrounded by commercial uses would more likely be developed into commercial use than into single-family residential use.

Five variables denoting the share of neighboring sites in each major land use (single-family, multi-family, commercial, industrial, and transportation) were included in each model. We computed these variables in two steps. First, for every site, we used GIS to identify the (initial) land uses of each adjacent site. (As noted above, sites are represented by 100m by 100m grid-cells. Thus, every grid-cell is surrounded by eight other grid cells.) Next, we computed the initial percentage of surrounding sites (or grid cells) in residential use, commercial use, industrial use, public use, or being used for transportation facilities. The resulting percentages vary between 1 and 0: a value of 1 indicates that a site is completely surrounded by a particular use; a value of 0 indicates no level of adjacency.

Except for transportation uses, we would expect the estimated coefficients for each of these variables to be positive. That is, we would expect that the probability of a vacant site being developed would be higher if it were entirely surrounded by residential uses than if it were surrounded by a mixture of residential, commercial, industrial, and public uses. The same logic would also apply for other developed land uses.

It would not necessarily apply to for surrounding vacant sites. To the extent that vacant sites function as competition for other vacant sites, the higher the percentage of surrounding vacant sites, the higher the level of competition, and thus the lower the likelihood that a particular vacant site might be developed.

The opposite interpretation is also possible. To the extent that there are minimum size thresholds for particular types of development, developers would tend to favor vacant sites surrounded by other vacant sites. To determine which effect dominates, we used GIS to count the number of vacant hectare grid-cells within 200 meters of each site. A negative coefficient would indicate that the competitive-supply effect dominates the size-threshold effect; a positive coefficient would indicate the opposite.

6b.<u>Inter-Use Externalities and Proximity Effects</u>: Different land uses generate both positive and negative externalities. Indeed, the desire to mitigate negative externalities has long been the classic argument behind zoning. Industrial uses are typically separated from residential uses to minimize aesthetic, safety, and property value spillovers. Likewise, high-density residential development is often separated from lower-density single-family development to minimize the potential for noise, traffic, and other potential spillovers.

Externalities need not be positive. The shopping center located near a large subdivision or apartment building is the beneficiary of a positive spillover--in this case, a large potential market. Similarly, the current emphasis being given mixed-use development is based on the presumption that for certain mixtures of uses, positive externalities exceed negative externalities.

To test for the existence and importance of inter-use externalities, we used GIS to measure the euclidean distance from each site to the nearest commercial site, industrial site, and public use site. To the extent that the negative externalities associated with a combination of uses exceed the positive externalities, we would expect the estimated coefficients of these variables to be negative. Conversely, a positive coefficient would indicate that the positive externalities associated with proximity between uses exceed the negative externalities. Finally, to the extent that proximity effects and inter-use externalities don't matter, we would expect the variable coefficients not to be statistically significant.

Table 2summarizes the various independent variables and their expected signs.

Table 2

Measuring Goodness-of-Fit

We begin by looking at how well the various models explain what actually happened--that is, by looking at the "goodness-of-fit" between the model predictions and actual land use changes. Unlike the r-square measure generated by linear regression, multi-nomial logit procedures do not generate a single goodness-of-fit measure. Instead, determinations of logit model quality are made according to whether the estimated models*correctlyclassify*the observations by category when compared with the observed category distribution. This idea of correct classification is often expressed in terms of the ratio of concordant predictions (those in which the predicted choice or change matches the observed choice or change) to the total number of observations. The number of concordant predictions can be determined using two different methods:

1.Maximum-Probability: The maximum-probability method is based on the assumption that the predicted choice or change is the one which has the highest probability of occurring regardless of the total number of observed choices or changes.

To illustrate the maximum-probability method and its problems, consider the case of an initially vacant site developed into residential use. (In terms of model calibration,*the observed change*in site status is vacant-to-residential.) The predicted change is estimated based on the results of the multi-nomial logit model. The predicted probabilities of land use change according to the model are .4 for conversion to residential use, .2 for conversion to commercial use, .1 for conversion to industrial use, and .3 that the site remains vacant. (Note that estimated probabilities sum to one.) Assuming the prediction process is unconstrained, the predicted change in site status will be from*vacant-to-residential*, corresponding to the highest change or choice probability. Note that the highest change/choice probability in this example is only .4.

Note that this method of case-by-case classification according to maximum probability is independent of the total number of observed choices. Suppose, for example, that the maximum-probability procedure described above were to generate 50 predictions of vacant-to-residential land use change, but that the actual number of observed vacant-to-residential land use changes was only 25. Depending on the individual model, the maximum probability method can significantly over- or underestimate the total number of observed category choices or changes.

2.Case-constrained: The case-constrained method imposes the additional assumption that the total number of predicted choices or changes can not exceed the observed number. In the example described above, only the first 25 vacant-to-residential land use changes would be counted as concordant.

The case-constrained method uses the maximum-probability measure to classify observations subject to the total case constraint. Once this constraint is met, an observation could potentially be classified into a lower-probability category. If for example, the total vacant-to-residential case constraint had already been met, the observation described above would be classified according to the next-highest probability--in which case it would remain vacant. The case-constrained method typically results in fewer total mis-predictions than the maximum-probability method, but is more likely to over-predict or under-predict in one or two categories.

Table 3presents both types of goodness-of-fit measures for all 16 multi-nomial land use change models, including eight vacant land models (one for each Bay Area County, except San Francisco) and eight redevelopment models.

Overall, the vacant land use change models fit the data reasonably well. Using the more reliable case-constrained good ness-of-fit-measure, overall model fits very from a high of 99.0% for Napa County (meaning that the model correctly predicted 99% of all observed 1985-95 grid-cell land use changes) to a low of 92.2% for Contra Costa County. Model fits using the maximum-probability method are comparable.

As a group, the models also do quite well at predicting which vacant sites are most likely to remain vacant. Case-constrained concordancy measures for "vacant-to-vacant" land use changes vary from a high of 99.5% for Napa County (meaning that the model correctly predicts that 99.5% of vacant sites in Napa County in 1985 would also be vacant in 1995) to a low of 96.0% in Contra Costa County. Given that most vacant sites in 1985 remained vacant through 1995 (seeTable 1), the very high level of concordancy for the vacant-to-vacant category is not surprising.

The various goodness-of-fit measures are much lower for vacant-to-single-family residential land use changes. They vary from a high of 48.3% in Solano (meaning that the model correctly predict only about half of vacant-to-residential land use changes) to a low of 21.6% in Sonoma

Table 3

County. Note that the maximum-probability goodness-of-fit measures are even lower than the case-constrained measures.

In the case of vacant-to-commercial land use changes, the two goodness-of-fit measures vary widely by county. In Napa County, the model correctly predicts 78.8% of vacant-to-commercial land use changes according to the case-constrained method, and 74.2% according to the maximum-probability method. In neighboring Contra Costa County, the same model specification correctly predicts only 12.4% of vacant-to-commercial land use changes using the case-constrained measure, and 0% using the maximum-probability measure.

There is also considerable goodness-of-fit variation between counties for vacant-to-industrial land use changes. The case-constrained measure ranges from a high of 47.9% in Napa County, to a low of only 6.1% in nearby Contra Costa County. Except for Marin, the other seven counties fall between these extremes.

Vacant-to-apartment land use changes were the least common form of major land use change in the Bay Area between 1985 and 1995 (seeTable 1). Because there were so few vacant-to-apartment observations upon which the various logit models could be calibrated, their ability to correctly predict such changes is limited. Prediction concordancy varies from a high of 31.3% in Solano County (using the case-constrained goodness-of-fit measure) to a low of only 10.4% in Contra Costa County. As with the other developed categories, note the generally poor quality of the maximum-probability goodness-of-fit measures.

The eight*redevelopment*models also explain overall rates and patterns of redevelopment--or more precisely, the lack of redevelopment--reasonably well. Using the more reliable case-constrained good ness-of-fit-measure, overall model fits very from a high of 95.8% for Solano and Sonoma counties, to a low of 91.4% in Santa Clara County. These high levels of model fit are, for the most part, based on how well the models explain the lack of redevelopment that occurred in the Bay Area between 1985 and 1995 (See Table 1). Case-constrained concordancy measures for "no change" in developed use vary from a high of 97.9% in Sonoma County, to a low of 95.5% in Santa Clara County.

The redevelopment models have much less ability to predict actual redevelopment. For residential redevelopment (that is, for sites that changed from a non-residential use in 1985 to a residential use in 1995), the case-constrained measure of concordancy varies from a high of 24.7% in Alameda County, to a low of 14.7% % in Santa Clara County. The ability of the models to explain commercial redevelopment (sites which changed from a non-commercial use in 1985 to a commercial use in 1995) is no better: case-constrained concordancy measures vary from a high of 49.7% in Alameda County, to a low of 2.6% in Napa County. The redevelopment models do a somewhat better job explaining industrial redevelopment (sites which changed from a non-industrial use in 1985 to an industrial use in 1995). Model fits for industrial redevelopment range from 66.7% in San Mateo County (meaning that the models correctly predict 66.7% of industrial redevelopment cases), to a low of 14.1% in Santa Clara County.

Models of Vacant Land Change: Parameter Estimates and Significance

Multinomial logit models generate much more and much richer output than conventional regression models. Indeed, for complicated models such as these, the amount of output can sometimes seem overwhelming. Rather than present the various model results*by land use within county*(the form in which they were originally calibrated), we present them*by county within use*. This makes it possible to compare the role and importance of different explanatory factors between counties.

Conversion of Undeveloped Sites to Single-family Land Uses:Residential development accounted for almost three-quarters of Bay Area vacant land development between 1985 and 1995, so the robust reliability of the logit models which capture those changes is not surprising (Table 4). As expected, and regardless of the county, demand factors played a consistent role in explaining vacant-to-single-family land conversion.

The effects of individual demand factors, however, were not always as predicted. Contrary to expectations, the higher the rate of job growth in a city between 1980 and 1985,*the lower the*

Table 4

*probability*that an undeveloped site within that city would be developed in single-family use. Household growth rates (between 1980 and 1985) were positively correlated with 1985-95 vacant-to-single-family land use conversion in San Mateo, Santa Clara, Solano, and Sonoma counties--as expected--but not in Alameda, Contra Costa, Marin, and Napa counties.

The effect of city size on residential land conversion also varied by county. Vacant sites in cities with larger populations in Alameda, Santa Clara, and Solano Counties were slightly more likely to be developed in residential use; while residential development favored sites in smaller cities in Contra Costa and San Mateo counties. Vacant sites were more likely to be developed into single-family residential use in cities which were job centers in 1980 in Alameda, Contra Costa, San Mateo, and Santa Clara Counties, but less likely to be developed into residential use in employment-rich cities in Napa and Solano counties. The general preference of residential development for job-rich cities is also evident from the positive coefficients associated with local jobs-housing ratios (although again, the sign, magnitude, and significance of this relationship can be seen to vary widely by county).

The rise of the edge cities and the forces of decentralization notwithstanding, vacant sites closer to San Francisco and/or San Jose were generally more likely to be converted into single-family residential use than more distance sites (as indicated by the negative coefficient estimates.) The importance of regional accessibility varied also widely by county. In Alameda County, for example, vacant sites closer to San Francisco or San Jose were neither more nor less likely to be developed into single-family use. Vacant sites in Contra Costa, Napa, and Solano counties, by contrast, were more likely to be developed in single-family use the closer they were to San Jose, and the farther they were from San Francisco.

The role of freeway proximity as a determinant of vacant-to-single-family land conversion also varied by county. Undeveloped sites close to freeways in Alameda, Contra Costa, Marin, Solano, and Sonoma counties were more likely to be developed into single-family use, while similar sites in Napa, San Mateo, and Santa Clara counties were less likely to be developed into single-family use. Likewise, proximity to a BART station served to discourage single-family residential development in both Alameda and Contra Costa counties.

Except in Alameda County, steeply sloped sites were far less likely to be developed into single-family use than flatter sites. The higher probability of residential development associated with moderately steep sites in Alameda County was probably due to the fact that those sites commanded bay and valley views.

All else being equal, prime farm sites were far less likely to be converted to single-family uses than other under undeveloped sites. this was especially true in Contra Costa and Solano counties. (Whether this disassociation between residential development and prime farm status is the result of local land use controls, or the ability of farmland to remain a viable economic use, or both, can not be determined from the model coefficients.)

The ability of sphere-of-influence boundaries to function as defacto urban limit lines between 1985 and 1995 was virtually nonexistent. In every county except for Contra Costa, vacant sites located inside existing spheres of influence were far less likely to be developed than comparable site outside sphere boundaries.

The extent to which distance from the urban edge served to discourage so-called "leapfrog" residential development also varied by county. All else being equal, the further an Alameda, Napa, or Solano site was outside an existing sphere-of-influence boundary, the more likely it was to be developed in residential use. By contrast, sites in Contra Costa, San Mateo, and Sonoma counties far beyond sphere-of-influence boundaries were less likely to be developed in single-family use than closer-in sites.

Vacant sites surrounded by residential uses were no more nor less likely to be converted to residential use that sites surrounded by other uses, or by a mixture of uses. Vacant sites surrounded by commercial or industrial uses, however, were generally less likely to be converted to single-family residential use. (In Contra Costa and Marin counties, by contrast, the higher the percentage of adjacent sites in commercial use, the greater the probability that a vacant site would be converted to residential use.)

Also inconsistent was the effect of vacant land supplies on undeveloped-to-residential land conversion. Vacant sites in Contra Costa, Marin, San Mateo, and Solano counties which were surrounded by other vacant sites were more likely to be converted to single-family use--suggesting that residential developers in these counties typically require larger sites. The opposite was true in Santa Clara County, were vacant sites were more likely to be developed into residential use if they were*not*surrounded by other vacant sites.

To the extent that there are any spillover-effects from commercial uses to residential uses, they are mostly positive. Regardless of county, the closer a vacant site was to a commercial use, the more likely it was to be developed in single-family residential use. As expected, the opposite was true for industrial uses: all else being equal, vacant sites near industrial uses were less likely to be developed in single-family use than other vacant sites. sites. (The two exceptions to this were in San Mateo and

Solano counties, where vacant sites near industrial uses were actually more lily to be developed in residential use.)

The extent to which public uses generated positive or negative externalities varied by county as well. In Alameda, Napa, Santa Clara, Solano, and Sonoma counties, vacant sites near existing public uses (typically schools and parks) were more likely to be developed in single-family residential use. In Contra Costa County, by contrast, vacant sites located near existing public uses were actually less likely to be developed in residential use.

Conversion of Undeveloped Sites to Multi-family Residential Uses:We begin our discussion of the results of the vacant-to-multifamily land use model (Table 5) with a caveat. Whether in terms of overall fit or coefficient reliability, the vacant-to-multi-family land use change model is less robust than the vacant-to-single-family model discussed above. This is partly because there were relatively few instances of vacant to-multi-family land use change between 1985 and 1995. It also follows from the fact that we estimated rather than directly observed the locations of sites in multi-family use in 1995. Altogether, we estimate that 263 hectares of undeveloped land were converted to multi-family use between 1985 and 1995. Vacant-to-multifamily land conversion accounted for a little more than one percent of Bay Area land use change between 1985 and 1995.

The effects of individual demand factors upon vacant-to-multi-family land use change varied sharply between counties. In Alameda County, for example, vacant sites located in larger cities and cities that had experienced recent (1980-85) household growth were more likely to be developed into multi-family residential use. In Contra Costa County, by contrast, where vacant-to-multi-family land conversion was concentrated in cities which were job-rich, but not necessarily large in terms of population. In Santa Clara County, vacant sites were more likely to be developed in residential use if they were located in cities that had experienced significant prior household growth, but not job growth. In Solano County, multi-family development was

Table 5

concentrated in larger cities, while in Sonoma County, it was focused in smaller cities.

The effect of regional accessibility on vacant-to-multi-family land conversion was far more consistent. Except for San Mateo and Solano counties, the closer a vacant site was to San Francisco, the more likely it was to be developed in multi-family use. Except for the North Bay counties, proximity to San Jose was also strongly correlated with multi-family land conversion.

The role of freeway proximity as a determinant of vacant-to-multi-family land conversion was much less consistent. Undeveloped sites close to freeways in Marin, Santa Clara and Solano counties were more likely to be developed into multi-family use, while similar sites in Contra Costa and San Mateo counties were less likely to be developed into multi-family use. Proximity to a BART station also served to discourage vacant-to-multi-family conversion in Contra Costa County; in Alameda County, proximity to a BART station had no significant effect--positive or negative--on vacant-to-multi-family land conversion.

Except in Marin County, steeply sloped sites were far less likely to be developed into multifamily use than flatter sites. Indeed, in most counties, the minimal number of vacant-to-multi-family land use conversions on anything other than flat sites made it difficult for the model to converge on a unique coefficient estimate. The same was also true for undeveloped prime farmland sites, and for vacant sites located inside city spheres-of-influence. Somewhat curiously, vacant sites in Contra Costa, Napa, San Mateo, Solano, and Sonoma counties were somewhat more likely to be developed in multi-family use the farther they were outside a sphere-of-influence boundary.

No discernable association is evident in any Bay Area County between the probability that a vacant sites would be developed in multi-family use, and the type and/or mixture of surrounding land uses. Proximity to an existing commercial use, by contrast, was associated with a higher probability of multi-family residential land development in every Bay Area county except Napa and Sonoma In a somewhat surprising finding, vacant sites near existing industrial uses were more likely to be converted multi-family use in Contra Costa, Marin, Napa, Solano, and Sonoma counties. Vacant sites near schools, parks, and other public uses were more likely to be developed into multi-family use in Alameda, Santa Clara, and Solano counties, but less likely to be developed into multi-family use in Contra Costa, Napa, and Sonoma counties.

Conversion of Undeveloped Sites to Commercial Uses:The commercial land use category subsumes all measure of retail, service, and office land uses. The 1980s was a period of tremendous commercial construction throughout the Bay Area, and new commercial development accounted for 14.7% percent of region wide vacant conversion between 1985 and 1995.

All else being equal, one might expect commercial development to be more commonplace in job-rich communities, or in fast-growing job centers. AsTable 6shows, this was indeed the case for vacant sites in Alameda and Contra Costa counties. In San Mateo, Santa Clara and Solano counties, however, commercial development favored sites in cities with slower-growing economies, and/or in cities with fewer jobs. (There were too few cases of vacant-to-commercial development in Martin and Napa counties upon which to calibrate a logit model.)

The commercial land use category includes retail centers. Accordingly, one might expect to find somewhat higher rates of vacant-to-commercial land use change among sites in cities with larger and/or fast-growing populations. This was indeed the case in San Mateo, Santa Clara, Solano counties, and to a much lesser extent, in Sonoma County. It was not the case in either Alameda or Contra Costa Counties, were vacant-to-commercial land use change seems to have favored sites in smaller and/or slower-growing cities.

New commercial development in the Bay Area remains very much oriented toward San Francisco. All else being equal, vacant sites closer to San Francisco were more likely to be converted to developed in commercial use than more distant sites. (Napa County sites were an exception to this.) Accessibility to San Jose--the Bay Area's other major job center--was less consistently associated with the probability that a vacant site would be commercially developed.

Freeway accessibility is often thought to be the single-most important factor influencing commercial development. Surprisingly, the role of freeway proximity as a determinant of vacant-to-commercial land conversion varied widely. In Alameda, Contra Costa, Marin, and Solano counties, undeveloped sites near freeway interchanges were far more likely to be developed in commercial use than mor distant sites. In San Mateo, Santa Clara, and Sonoma counties, by contrast, proximity to a freeway interchange did not seem to play a significant role in encouraging new commercial development. Napa County was again an outlier, as the likelihood that vacant sites would be developed in commercial use actually increased with

<u>Table 6</u>

distance from a freeway interchange.

As was also the case for residential development, steeply sloped sites were far less likely to be commercial developed than flatter sites. Vacant sites in Alameda, Contra Costa, and Santa Clara counties were also much likely to be developed in commercial use if they were located on prime farmland. In Marin, Napa, San Mateo, Solano, and Sonoma counties, by contrast, a particular site's farmland status did not seem to affect its commercial development potential. affect

Regardless of county, the ability of sphere-of-influence boundaries to contain new commercial development appears to have negligible. All else being equal, vacant sites outside existing sphere-of-influence boundaries were more likely to have been developed in commercial use than vacant sites inside sphere-of-influence boundaries.

The extent to which distance from the urban edge served to discourage commercial "leapfrogging" also varied widely by county. All else being equal, the further a site in Alameda, Marin, Santa Clara, or Solano counties was outside an existing sphere-of-influence boundary, the higher its probability of development. The opposite was true in Contra Costa, Napa, San Mateo, and Sonoma counties, where sites closer to sphere-of-influence boundaries were more likely to be commercially developed.

In general, vacant sites surrounded by residential uses were less likely to be developed into commercial use than sites surrounded by other uses, or by a mixture of uses. (Here again, Napa county sites proved an exception to this generalization) . Vacant sites surrounded by other vacant sites were more likely to be developed into commercial use, suggesting the importance of large-scale land assembly to commercial development. As above, the significance of this effect varied widely by county.

Commercial land uses are subject to both centripetal (clustering) and centrifugal (dispersing) forces. In general, the clustering forces are stronger. Vacant sites in Alameda, Contra Costa, San Mateo, Solano, and Sonoma counties were much more likely to be developed in commercial use the closer they were to other existing commercial land uses. (In Marin and Napa counties, by contrast, the probability that a vacant site would be developed in commercial use decreased with proximity to existing commercial land uses.. New commercial developments , like new residential developments, were more likely to be repelled by than attracted to existing industrial land uses. (Once again, commercial development in Napa County proved to be a consistent exception to this generalization.)

Conversion of Undeveloped Sites to Industrial Uses:The industrial land use category subsumes all manufacturing and production facilities as well as warehouse, distribution, and construction uses. Altogether, new industrial development accounted for 6.3% of Bay Area vacant land change between 1985 and 1995.

The factors which explain industrial development patterns in the Bay Area differ widely by county (Table 7). In Alameda County, for example, a vacant site was more likely to be developed into industrial use if was flat, near a freeway interchange, or near an existing commercial or industrial land use, or outside an existing sphere-of-influence boundary. Vacant Alameda sites surrounded by residential or commercial uses were less likely to be developed in industrial use. Vacant sites in job-rich cities, or cities that experienced recent job growth were also less likely to be industrially developed. A similar set of factors accounted for industrial development patterns in Santa Clara County. Additionally, Santa Clara sites were less likely to be converted to industrial uses if they were on prime agricultural land.

Two other counties, Solano and Sonoma, experienced significant amounts of new industrial development between 1985 and 1995. New industrial development in Solano County was attracted to existing industrial sites, but repelled from commercial sites. Sloped sites, sites located far beyond sphere of influence boundaries, and prime agricultural land sites were also less likely to be developed in industrial use. Curiously, Solano County sites located near freeway interchanges were less, not more likely to be developed in industrial use. As in Alameda and Santa Clara counties, new industrial development favored sites in or near cities with growing populations, but not expanding economies. In Sonoma County, new industrial development followed still a different pattern, favoring flat sites in the southern part of county, sites near existing industrial uses, and sites beyond city and sphere-of-influence boundaries.

Table 7

Parameter Estimates and Significance: Redevelopment Models

Redevelopment is typically a much more complicated and idiosyncratic process than new development. Whereas new development creates (or at least helps create) the urban fabric, redevelopment typically occurs within the context of an existing urban fabric. This drastically increases the number of interests or stakeholders involved in the redevelopment process. It also increases the uncertainty associated with a particular redevelopment project. Getting the politics right is often more important to successful redevelopment projects then getting the economics right. Because local politics are notoriously difficult to model, the statistics (including both coefficient estimates and measures of goodness-of-fit) associated with the various redevelopment models tend to be much more uneven that the statistics associated with the new development models, above.

Redevelopment to Residential Use: The usual image of urban redevelopment is that it is a process whereby older residential or industrial sites are recycled into newer commercial and business uses. In fact, redevelopment*to*residential use dominated all other forms of redevelopment in the Bay Area between 1985 and 1995. According to the Association of Bay Area Governments, more than 7,800 hectares of land were redeveloped into single-family or apartment use between 1985 and 1995.

Despite its apparent popularity, the factors behind residential redevelopment in the Bay Area are far from systematic. In fact, asTable 8shows, the factors which explain residential development differ widely and erratically between counties. Indeed, the only factor which consistently explains recent patterns of residential redevelopment in more than one county was distance to a sphere-of-influence line. Except in Alameda County, the closer a particular site was to a sphere-of-influence line, the more likely it was to be redeveloped into residential use. Contrary to the new development case, above, sites near freeway interchanges were less likely, not more likely, to be redeveloped into residential use

All else being equal, residential redevelopment in Alameda County favored sites in the southern part of the county, as well as sites in larger cities. Sites near freeway interchanges or BART stations, or in job-rich cities were less likely to be redeveloped into residential use. In Contra Costa County, residential redevelopment favored sites close to San Francisco, as well sites near existing commercial land uses. The same two factors contributed to residential

Table 8

redevelopment in Santa Clara County, as did local rates of household growth, and nearby vacant land.

Santa Clara County sites less likely to be redeveloped into residential use were those located near freeway interchanges, those with any type of slope, and those located in job-rich cities. In Marin County, residential redevelopment was (and is) so rare that none of the factors included as independent variables could explain it. In Napa County, residential redevelopment between 1985 and 1995 favored sites far away from freeways as well as sites near existing commercial and industrial uses. Proximity to industrial uses also served to encourage residential redevelopment in San Mateo County, although curiously, proximity to commercial uses had the opposite effect. In Solano County, sites near industrial uses were more likely to be redeveloped in residential use, although sites surrounded by industrial land uses were not. Sites in Sonoma County were more likely to be residentially redeveloped the closer they were to freeways and industrial uses, but the further they were from residential uses.

In almost no Bay Area County did a particular site's initial land use, its proximity to public land uses such as parks or schools, or the mix of adjacent land uses affect the likelihood that it would be residentially redeveloped.

Redevelopment to Commercial Use: Four Bay Area counties--Alameda, Contra Costa, Santa Clara, and Sonoma--experienced significant amounts of commercial redevelopment (i.e., land use change from a developed non-commercial use to a commercial use) between 1985 and 1995. Commercial redevelopment was an extremely infrequent form of land use change in the region's other counties.

Commercial redevelopment is as idiosyncratic and difficult to model as residential redevelopment. Of the many explanatory factors listed inTable 9only two were consistently associated with higher rates of commercial redevelopment between 1985 and 1995: initial proximity to a commercial use, and municipal household growth.

The effects of other explanatory factors varied by county. All else being equal, commercial redevelopment in Alameda County favored flat sites closer to San Francisco, or near freeway interchanges, or surrounded--at least initially--by residential uses. In neighboring Contra Costa County, commercial redevelopment favored flat sites far from public uses, and/or sites located in

Table 9

cites with declining economies such as Richmond. In Santa Clara County, the sites most likely to be commercially redeveloped between 1985 and 1995 were those located in existing commercial areas, or those close to downtown San Jose, or those initially in industrial or transportation use. Santa Clara County was the only Bay Area county in which a site's initial use played any role in shaping whether it was likely to be commercially redeveloped. In Sonoma County, commercial redevelopment favored sites close to freeway interchanges, sites closer to San Francisco, and sites within or close to sphere-of-influence boundary lines.

Redevelopment to Industrial Uses: Industrial redevelopment--that is, redevelopment from a residential, commercial, or public use to an industrial use--is an infrequent occurrence in the San Francisco Bay Area. (Most industrial development occurs on previously undeveloped land.) Region wide, just over 200 hectares of urban land were redeveloped to industrial use between 1985 and 1995.

The majority of the region's industrial redevelopment occurred in just two counties: Santa Clara and Sonoma (Table 10). In Santa Clara, the sites most likely to be industrially redeveloped between 1985 and 1995 were those which were close to freeway interchanges, those which were close to existing commercial or industrial uses, those which were inside sphere-of-influence boundaries, and those in large but not necessarily job rich cities. Proximity to San Francisco, but not necessarily San Jose also contributed to the likelihood that a site would be redeveloped into industrial use.

The pattern of industrial redevelopment in Sonoma County was somewhat different. Sonoma County sites most likely to be industrially redeveloped included those near freeways, those near existing commercial uses, and those outside city or sphere-of-influence boundaries. Industrial redevelopment in Sonoma County was concentrated in sites in cities with expanding economies but not necessarily expanding populations.

The counties with the next largest totals of industrial redevelopment between 1985 and 1995 were Alameda and Napa. Industrial redevelopment in Alameda County favored flat sites, sites in larger cities, sites close to BART stations and sites close to commercial uses. Proximity to freeway interchanges and other industrial sites, two factors which were extremely important in Santa Clara, were not important in Alameda County. In Napa County, industrial redevelopment was concentrated in sites near or adjacent to industrial land.

Table 10

SUMMARY AND CONCLUSIONS

The CUF II Model is a significant improvement over its predecessor and step forward in the continuing evolution of urban growth models. The foremost feature of the CUF II Model is that it allows competing land uses to bid against each other on a site-by-site basis, without any*a priori*assumption regarding which uses are superior or inferior. This is achieved by using a multi-nomial logit estimator for model calibration. The fact that the model is actually calibrated against observed site-level land use changes, as opposed to zone-based changes in activity levels, is also significant. To our knowledge, the CUF II Model is the first urban growth model to realistically incorporate the potential for urban redevelopment--that is, that a site might be redeveloped from one urban use to another.

Through its links to GIS, the CUF II Model makes it possible to identify the determinants of land use change at different spatial scales. The land use change model, for example, incorporates site-level measures such as slope and proximity to major transportation infrastructure; neighborhood effects such as the role of adjoining land uses in encouraging or discouraging development; city-scale effects such as population and employment growth; and regional effects such a accessibility to major employment centers. The land use change model also includes key policy variables such as site distance to the nearest sphere-of-influence boundary, and distance to the nearest rail transit station. The presence of all of these variables makes it possible for the CUF II Model to test the spatial impacts of an unprecedented set of realistic policy variables, including, for the first time, new freeway and transit facilities.

The CUF Model is not just more advanced that its predecessor, it is also different. Whereas the previous version of the CUF Model approximated sites as spatially homogeneous collections of development attributes (termed Developable Land Units, or DLUs), the CUF II Model represents potential development sites as hectare grid-cells. The shift to grid cells (made possible by advances in

GIS technology) make it possible for CUF II Model to incorporate multiple adjacency and nearest neighbor effects, as well as better measure distances.

Like its predecessor, the CUF II Model is modular. Alternative regional and/or local household and employment projections can be inserted into the model. The model can also be updated as new information becomes available. Perhaps most important, the CUF II Model is capable of simulating and summarizing processes of urban growth at the site level, just like its predecessor.

The Things that Matter

The CUF II Model shines a spotlight on the mechanisms and processes of urban land use change. In doing so, it brings to the forefront those factors which most matter. It also makes clear the interactions between the determinants of land use change and land use policy.

What causes a particular vacant site to be developed, or an already-developed site to be redeveloped? The answers to this question vary by place and land use. The factors most consistently associated with recent vacant-to-single-family residential land use change in the Bay Area have been: (i) proximity to a freeway interchange; (ii) low rates of community job growth, and; (iii) proximity to nearby retail activities. The availability of flat sites was found to important in every Bay Area county but Contra Costa. All else being equal vacant sites atop prime agricultural land were less attractive to residential developers than lower-quality agricultural sites. Infill sites (e.g., vacant sites within existing city spheres-of-influence) were generally less attractive than more far-flung sites. Sites near existing industrial uses, were, not surprisingly, less attractive to single-family home developers than other sites. Conversely, vacant sites near public uses, especially schools and parks, were more likely to be residentially developed. The closer a site was to downtown San Jose, the more likely it was to be developed to single-family use. Except in Sonoma and San Mateo counties, the same was not true for sites near downtown San Francisco. New housing development was slightly more likely to occur in cities with large numbers of jobs.

The same factors that explain single-family housing development patterns--freeway proximity, site slope, distance to downtown San Jose, and proximity to commercial uses--also explain apartment development patterns. New apartments, like new single-family homes, were not particularly attracted to infill sites. Nor were they consistently attracted to sites near public uses such as schools and parks. In Alameda County, proximity to BART had no effect on apartment development; in Contra Costa County, vacant sites near BART stations were less likely to be developed in apartment use.

Patterns of vacant-to-commercial and vacant-to industrial land use change also varied widely by county. New retail uses in Alameda and Contra Costa Counties, for example, favored cities with large numbers of jobs, sites closer to downtown San Francisco and San Jose, sites near freeway interchanges, and sites near other commercial uses. New retail uses in Santa Clara County, by contrast, favored sites in communities with fewer jobs and more houses, site closer to San Francisco but not downtown San Jose, and sites near existing industrial areas; they did not favor sites near freeways or other retail uses. In Marin County, new retail users preferred sites with direct freeway access and sites close to downtown, but were "repelled" by existing commercial land uses. Industrial users generally favored vacant sites in growing communities, sites with freeway access, and most of all, sites near other industrial users.

Redevelopment is typically a much more complicated and idiosyncratic process than new development. Accordingly, Bay Area patterns of redevelopment varied widely by terminal use and by county. Still, a few generalizations regarding redevelopment do apply. Redevelopment occurs only on flat or near-flat sites. Contrary to the new development case, freeway proximity was not a consistent determinant of redevelopment activity. Nor was proximity to a BART station. Except for industrial redevelopment, a site's initial use had little impact on its terminal use. Sites were more likely to be redeveloped to commercial and industrial uses the closer they were to other commercial and industrial uses (respectively). All else being equal, redevelopment activity in the Bay Area favored larger (but not necessarily job-rich) cities with growing populations but not necessarily growing economies.

All of these results are subject to four caveats. The first is that they are based on observed land use changes which occurred between 1985 and 1995. Their applicability to earlier periods, or for that matter, to the current one, is unclear. Second, these results are based on analyses of changes in dominant land use at the hectare, not parcel, level. To illustrate the limitations of this type of analysis, consider the case of a hectare grid-cell that, as of 1985, was 40 percent undeveloped, 30 percent residentially-developed, and 30 percent commercially-developed. Next, suppose that a single new home was built in 1987 on a vacant parcel within the same grid-cell. According to our (and ABAG's) definitions, the dominant land use of this grid-cell would have changed from 'undeveloped' in 1985, to 'residential' in 1995. These types of threshold-changes, while uncommon throughout the entire data set, were not unknown either. Third, while the various models do a good job predicting which sites*will not change*land use (e.g., remain vacant, or not be redeveloped), they are generally less capable of predicting which specific sites*will shift*land use. Clearly, there are many other factors that shape land use change which should be included in the various models. Fourth, the extent to which the model results reflect underlying public policy regimes, particularly local zoning, is unclear. The fact that new commercial development favored sites near freeways may simply reflect the fact that so many parcels near freeways are zoned for commercial development. Similarly, the observation that new residential development did not favor "prime" agricultural sites may be more indicative of the success of agricultural preservation policies than of underlying developer preferences.

An Agenda for CUF II.1 and for CUF III

Like all urban growth models, the CUF II Model still has plenty of room for improvement. The model is horrifically data hungry--although recent advances in GIS technology make that hunger easier than ever to satisfy. Despite its relatively simple structure, the model is not easy to understand or explain.

Its bid-based approach to allocating competing land uses, while intuitively appealing, sometimes leads to unrealistic results.

Judging from their middling goodness-of-fit results, the current multi-nomial logit land use change models are still incomplete. Factors like local zoning designations and codes need to be tested and added to the models, as do local public service quality and availability, and better measures or sub-regional accessibility. The extent of bias due to spatial autocorrelation also needs to be carefully investigated. The fact that the estimated logit equations are based on *all* sites rather than a random sample also introduces potential bias.

The biggest remaining gap in the CUF II Model is that it doesn't incorporate actual or imputed real estate prices. Prices serve three essential functions in urban real estate markets. The first is to provide independent signals to both buyers and producers regarding appropriate land uses and land use intensities. Second, and of greater importance, prices function to "clear" real estate markets. Market clearing is the process whereby individual transactions between willing buyers and sellers lead to marketwide outcomes that are fully reflective of supplier costs, and of demander preferences and incomes. Third, prices are the mechanisms through which markets adjust to short-term imbalances between supply and demand.

The CUF II Models's lack of prices is especially problematic when simulating spillovers. As currently structured, activities which can not be accommodated in one location (usually for reasons of insufficient land or because of policy constraints) costlessly spillover to the next best location. Within a particular radius (as determined by the user), there is no particular penalty associated with spillover. Real life doesn't work that way. Confronted with the possibility that their preferred sites may be unavailable, many activities will not consider alternative sites. They will instead raise their bid prices for their preferred sites. The result will be an increase in site prices (in the form of economic rents) together with some degree of spillover (by those activities which *will* consider alternative sites). Because the CUF II Model does not include explicit prices, it can not deal with different price elasticities of demand, or, for that matter, of supply. Its sole response to unaccommodated demand is through the spillover mechanism.

A second and related limitation of the CUF II Model is that it lacks agents. The multi-nomial logit models that form the heart of the CUF II Model are all reduced-form models. That is, they focus on the characteristics of transactions (land use change in this case) but give short shrift to the characteristics and motivations buyers (e.g., households and businesses) or sellers (e.g., land owners and developers). To the extent that the timing and nature of actual land use changes reflect the economic characteristics and personal motivations of real people and real businesses, and not just the locational characteristics of sites, the CUF II model may be regarded is seriously incomplete, or potentially biased, or both.

In the absence of agents, one must assume that all site demanders (households and businesses), and all site providers (developers and land owners) are functionally the same--which is to say that they value particular sites in the same ways. This assumption is neither true nor defensible. Households come in many different sizes, configurations, and lifecycles, and with sharply varying preferences for location, housing, travel, and public services. Businesses likewise are extremely diverse, even within a single industry. On the supply side, landowners often sell to developers for lifecycle rather than purely economic reasons. Likewise, the actual projects developers build often have more to do with what the entitlements process will allow than what the market will reward. Adding information on agents to the CUF Model should improve both its reliability and robustness.

How might such information be added? Perhaps the simplest approach would be to include the demographic and/or economic characteristics of local households and/or businesses as independent variables in the logit specifications. The simplicity of this approach notwithstanding, it would likely exacerbate problems of ecological fallacy. A better long-term approach would be to employ techniques of micro-simulation; that is to observe and then try to model the bid and acceptance functions associated with a representative sample of households, business, landowners, and developers.

Ultimately, of course, the logit equations used to explain land use change must be linked back to the regressions equations used to predict population and job growth. While is convenient to model land use change and activity allocation (or land use change) as separate phenomena, it is certainly not correct. Where new development of different types goes (and where it does not go); the price structure of new development; the extent to which development is imported or exported from different communities; the importance of infrastructure investments and inter-use externalities--all of these outcomes ultimately become the inputs to yet a new round of metropolitan-scale growth and change.

References

Association of Bay Area Governments (ABAG). 1995.*Bay Area Land Use Inventory.*Oakland: Association of Bay Area Governments.

Batty, Michael. 1992. Urban Modeling in Computer-Graphic and Geographic Information System Environments.*Environment and Planning B*19: 689-708.

Batty, Michael and Y Xie. 1994. From Cells to Cities.*Environment and Planning B: Planning and Design*21: 531-48.

Ben-Akiva M. and S. R. Lerman. 1985.*Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge: MIT Press.

Clarke, Keith.

Domenich, Thomas A. and Daniel McFadden. 1975.*Urban Travel Demand: a Behavioral Analysis*. Amsterdam: North-Holland.

Fisher, M. Manfred and Peter Nijkamp. 1985. Developments in Explanatory Discrete Spatial Data and Choice Analysis.*Progress in Human Geography*9: 515-51. California Department of Resources. 1986. Farmland Mapping Project.

Landis, John. 1994a. The California Urban Futures Model: A New Generation of Metropolitan Simulation Models.*Environment and Planning B: Planning and Design*. 21: 399-420.

Landis, John D., and Ming Zhao. 1994b. Pilot Study of Solano and Sonoma Land Use and Development Policy Alternatives. Working Paper 618. Berkeley, CA: Institute of Urban and Regional Development.

Landis, John. 1995. Imagining Land Use Futures: Applying the California Urban Futures Model.*Journal of the American Planning Association*. 61:1. 438-457.

Lee, Douglas B. 1973. Requiem for Large-scale Models.*Journal of the American Institute of Planners*39: 163-178.

McFadden, Daniel. 1973. Conditional Logit Analysis of Quantitative Choice Behavior. in P. Zarembka (ed.)*Frontiers in Econometrics*. New York: Academic Press.

U.S. Census Bureau. 1995. 1994 (Post-Census) California TIGER Files.

Waddell, Paul and Vibhooti Shukla. 1993. Manufacturing Location in a Polycentric Urban Area: A Study in the Composition and Attractiveness of Employment Subcenters.*Urban Geography*14: 277-296.

Wegener, Michael. 1994. Operational Urban Models: State of the Art.*Journal of the American Planning Association*60:1. 17-30.

Wegener, Michael. 1995. Current and Future Land Use Models. Proceedings of the Travel Model Improvement Program, Land Use Modeling Conference, February 19-21, 1995. Washington D.C.: U.S. Departments of Transportation and Energy, and the Environmental Protection Agency.

An Open Geographic Modeling System

Thomas Maxwell, Robert Costanza

University of Maryland,

Institute for Ecological Economics

Box 38, Solomons, Md. 20688

Abstract

Developing the complex computer models that are necessary for effectively managing human affairs through the next century requires new infrastructure supporting high performance collaborative modeling.

In this paper we describe an ongoing program to develop 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 multiple 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 support of science, education, and policy making on many levels.

---

1. Introduction

We believe that effectively managing human affairs through the next century will require extremely complex and reliable computer models. Programs such as NASA's Mission to Planet Earth (MTPE) augmented by the development of the Open Geodata Interoperability Specification (OGIS) by the Open GIS Consortium have paved the way by providing an open framework for interoperable geodata access and processing. However, the crucial next step, the development of an open framework to support collabopprative dynamic modeling, has been all but ignored. This paper describes an ongoing program to develop this crucial infrastructure for open geographic modeling.

Generally speaking, a dynamic geographic simulation is a computer-based model that simulates the spatio-temporal dynamics of a landscape at scales ranging from a watershed to the globe. It is structured as a spatial grid/network of site-models, each representing the dynamics of a landscape component. A site-model incorporates multiple mathematical, logical, and stochastic process models created and defined by a multidisciplinary research team. These process models calculate the state of the system at a time t+dt as a function of the state at time t. The initial conditions are seeded by the use of a "picture" or "snapshot" of the system at some real or artificial start time. Seeding may use various forms of input data including raster images, vector data, point information, and object status and location. Recent advances in computer processing speed and efficiency have made it possible to incorporate spatial components of landscapes -i.e. population distribution, specific landscape conditions- into dynamic, geographic models.

Spatially explicit (geographic) modeling is essential for realistically addressing many resource management and environmental impact scenarios related to industry, agriculture and the natural environment. Across the nation, land management offices responsible for the management of natural resources, endangered species, water quality, aesthetics, and economic productivity of the land, are turning toward dynamic, geographic modeling. The development of these simulations has in general 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 complexities of science are outstripping the capability of any single institution to embrace the full range of pertinent expertise in any area" (Towards a National Collaboratory, NSF Workshop 3/17 '89). Collaborative, interpersonal computing must become a focus for efforts to increase the possibilities for and productivity of interdisciplinary teams working together on geographic modeling projects. This community must be tightly coupled within a worldwide "collaboratory" based on new electronic information sharing technologies, bringing together experts in computational and life sciences with policy makers and stake holders. The initial groundwork for this collaboratory has been laid with the development of the Open GIS interoperability standard (OGIS). Over the past six years our team has been developing strategies and software designs to take the next step toward a worldwide Open Geographic Modelling System (OGMS).

---

2. Supporting Collaborative Modeling

There exists a rich set of research problems associated with the implementation of computer based collaborative technologies for high performance geographic modeling. Five important areas of ongoing research and development are integrated support for 1) modular, collaborative model development, 2) transparent access to high performance computing resources, 3) graphical display & manipulation of model structure and dynamics, and 4) multiple integrated spatio-temporal representations.

2.1 Collaborative Modular Modeling

The development of team approaches to computer-based resource management reflects the increasing reliance on experts from multiple disciplines with the variety of specialties necessary to address all aspects of the complex management scenarios. The development of realistic models in this field can require collaboration between species specialists, hydrologists, chemists, land managers, economists, ecologists, and others.

Supporting collaborative modeling requires 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 design facilitates collaborative model construction, since teams of specialists can work independently on different modules with minimal risk of interference. Modules 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 archived object. A modeling environment that supports modularity could provide a universal modeling language to promote worldwide collaborative model development.

2.2 High Performance Computing

High performance computing gains its strength from transparent availability of computational resources. Developments to date have been focused on new hardware and software architectures designed to efficiently map an application to the computing environment, and methods of distributing processes to appropriate platforms. Much less effort has gone into efficiently mapping the teams of researchers who use these high performance systems into the high performance environment. For this reason high performance systems have seen little use in the field of environmental modeling and resource management, even though this class of models is a near ideal application for distributed processing (i.e. a typical model consists of a large number of cells, each containing a computationally intensive unit model that can be executed semi-independently; each computational node can be assigned a different subset of cells, and most inter--node communication is nearest-neighbor only). Our development program is decicated to addressing this issue by transparantly integrating high performance distributed computing resources into the model development environment.

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, debugging, 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 students, policy makers, scientists, 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 geographic 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 multiple 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 linking, transferring data between, and decomposing (over multiple processors) spatial representations should be transparent to the modelers.

2.5 Multiple Temporal Modes

Building realistic geographic 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 emulated 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 geographic model development, our team has been developing an integrated environment for high performance spatial modeling, called the Spatial Modeling Environment (SME). This environment, which transparently links icon-based modeling environments with advanced computing resources, allows users to develop models in a user-friendly, graphical environment, requiring very little knowledge of computers or computer programming. Automatic code generators construct spatial simulations and enable distributed processing over a network of parallel and serial computers, allowing transparent access to state-of-the-art computing facilities. The modeling environment imposes the constraints of modularity and hierarchy in program design, and supports archiving of reusable modules in our Modular Modeling Language (MML). An associated library of "module wrappers" will facilitate the incorporation of legacy simulation models into the environment. This paradigm encourages the development of libraries of modules representing model components that are globally available to model builders, enabling users to build on the work of others instead of starting from scratch each time a new model is initiated. A menu-driven run-time interface will provide the user with a single familiar environment in which to interact with simulations running on any one of a number of parallel or serial computers. This interface will allow users with widely varying goals and background knowledge (from scientists and students to policy makers) to easily control the simulation parameters and generate graphical output (at their Web browser) in a manner appropriate to their level of expertise. The adoption of this paradigm should greatly facilitate the application of computer modeling to the study of spatial systems in support of research, education, and policy making.

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 inclusive of the full range of relevant geographic modeling activities. 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 (SME). The three-part Modelbase-View-Driver architecture of the SME is 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 application such as STELLA (HPS) or Extend (Imagine That Inc., San Jose, CA).

3.2 Modelbase

The ModuleConstructor application converts the View ecosystem component modules into Module objects defined in our text-based Modular Modeling Language (MML). The MML objects can then be archived in the ModelBase to be accessed by other researchers, and/or used immediately to construct 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 drive a spatial simulation.

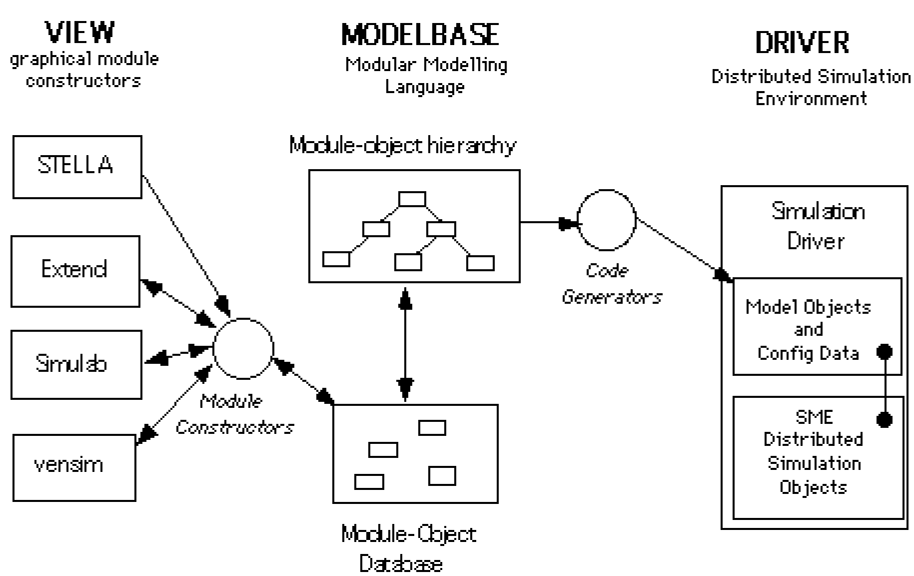


Figure 1: The ModelBase-View-Driver Architecture.

3.3 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 representations 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.

3.4 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 message passing, layered on top of the PointGrid library.

---

4. Current Applications

The current management applications of this framework include the Everglades Landscape Model (ELM), the University of Illinois GMSLab's Threatened and Endangered Species Models (TESM), and the Patuxent Landscape Model (PLM). The SME is also being used for educational purposes at the University of Illinois to support ecosystem modeling classes. These examples demonstrate the use of the SME by interdisciplinary research teams coordinating similar large-scale, ecological modeling efforts. Additional resource management and education programs will be initiated as the system moves out of the development stage.

4.1 Patuxent Landscape Model

The Patuxent Landscape Model (PLM) is 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 developed 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 contains about 6,000 spatial cells each containing a dynamic simulation model (based on the GEM model (Fitz et al., 1995)) containing approximately 20 state variables partitioned into 14 modules. It uses two frames, a 2D grid frame (for modules such as 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.

4.2 Everglades Landscape Model

The Everglades Landscape Model (ELM) is designed to be one of the principal tools in a systematic analysis of the varying options in managing the distribution of water and nutrients in the Everglades. This system has myriad indirect interactions, constraints and feedbacks that result in complex ecosystem structure (biotic and abiotic components and their flow pathways) and function (the modes of interaction and their rates). For this reason, it is critical to develop a systems viewpoint towards understanding the dynamics inherent in that ecosystem structure and function. Part of this process is the development of a dynamic spatial simulation model. Using this spatially explict process model, changing spatial patterns and processes in the Everglades landscape can be analyzed within the context of altered management strategies. Only by incorporating spatial articulation can an ecological model realistically address large scale management issues within the vast, heterogeneous system of the Everglades.

4.3 U.S. Military Applications

A unique example of an organization working to understand and manage complex landscape systems is the US military. Having sequestered large tracts of land to function as military training centers, the military is now responsible, as mandated by guidelines in the Endangered Species Act, for managing and preserving the vast collection of ecosystems living on those landscapes. In support of this effort a group of scientists and students at the University of Illinois have developed a number of dynamic spatial ecosystem models designed to help manage and protect a threatened species. The species modeled includes the desert tortoise (Gopherus agassizii), living at Fort Irwin, a U.S. Army training center in the central Mojave Desert of California. The Desert Tortoise model is a collection of four separate sub-models, all constructed within the SME modeling environment. Each model was built by separate groups of the research team members. This division of labor and cross-disciplinary approach allowed individuals to apply their expertise to the most appropriate aspects of the project. The final model is an interdependent web of four distinct aspects of Fort Irwin's landscape: climate, including hydrology and temperature; vegetation; tortoise population growth; and tortoise movement.

5. 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 advanced 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.

---

6. References

Acock, B. and Reynolds, J.F., 1990. Model Structure and Data Base Development. In: R.K. Dixon, R.S. Meldahl, G.A. Ruark and W.G. Warren (Editors), Process Modeling of Forest Growth Responses to Environmental Stress. Timber Press, Portland, OR.DMSTS, 1995.

Fitz, H.C., DeBellevue, E., Costanza, R., Boumans, R., Maxwell, T. and Wainger, L., 1995. Development of a General Ecosystem Model (GEM) for a range of scales and ecosystems. Ecological Modelling, 88: 263-297.

Gauthier, R.L. and Ponto, S.D., 1970. Designing Systems Programs. Prentice-Hall, Englewood Cliffs, NJ.Goodall, D.W., 1974.

Maxwell, T. and Costanza, R., 1994. Spatial Ecosystem Modeling in a Distributed Computational Environment. In: J. van den Bergh and J. van der Straaten (Editors), Toward Sustainable Development: Concepts, Methods, and Policy. Island Press, Washington, D.C.

Maxwell, T. and Costanza, R., 1995. Distributed Modular Spatial Ecosystem Modelling. International Journal of Computer Simulation: Special Issue on Advanced Simulation Methodologies, 5(3):247-262.

Maxwell, T. and Costanza, R., 1997. An Open Geographic Modeling Environment. Simulation, 68(3):175-185.

Silvert, W., 1993. Object-Oriented Ecosystem Modeling. Ecological Modeling, 68:91-118.

The Syntax and Semantics of Urban Modeling: Versions vs. Visions
[Draft, not for citation without the author's permission]

Daniel Z. Sui
Department of Geography
Texas A&M University
College Station, TX 77843-3147

1. Introduction

Originally developed to simulate the impacts of various policy initiatives for urban development during the post-war economic boom, urban modeling bloomed in the late 50s and throughout the 60s in both the U.S. and Western European countries. With the massive transformation from an industrial to an informational economy and the shift in political ideology, strong oppositions were leveled against urban modeling in the early 70s. Although several groups of urban modellers continued to carry out their research within the academia, urban modeling gradually faded away as a dominant planning and decision making paradigm in the late 70s and through most of the 1980s. The growing popularity of GIS and the increasing data collection efforts by both the public and private sectors have created new needs for urban modeling (Crawford-Tilley et al., 1996). We began to witness a revitalization of urban modeling since the late 80s. By the mid 90s, we saw a full-fledged renaissance of large-scale urban modeling as evidenced by the increasing integration of GIS with urban modeling via various loose and close coupling strategies (Sui, 1997b).

The objectives of this paper are two fold. First, I want to briefly trace the historical development of various versions of urban models. Second, I will present some thoughts on future urban modeling efforts. I understand both of these two tasks are enormously complex and challenging, and may be well beyond my ability to handle. What I am aiming at is that my ambitious goals will reach one humble end: to provide some food for thoughts and hopefully, to chart a better map for future urban modeling through the collective wisdom provided in this workshop.

I need a ladder to accomplish the twin goals of this paper. The ladder I will use is John Casti's conceptual framework for a science of surprise (Casti, 1994). In the second part of this paper, I will briefly introduce Casti's idea on the science of surprise. Then I will use Casti's framework to analyze the different versions of urban modeling and to discuss some new visions for our future modeling efforts.

2. Casti's Science of Surprise

John L. Casti, who is widely regarded as one of the most prolific mathematicians and science writers in the world, has discussed in several of his recent books the fundamental nature of modeling and the reasons why our models sometimes fail (Casti, 1991; 1994; 1997). According to Casti (1994), surprise occurs when the expected results of our models do not match the reality we are trying to predict. The science of modeling is actually a science of surprise. To understand modeling, we must understand the mechanisms that contribute to the generation of surprise and find out how to deal with them. Essentially, Casti's science of surprise consists of the following two components:

2.1. The nature of modeling:

According to Casti (1994), the essence of modeling is a two-way mapping process: to encode certain characterizations (observables) in a natural (real world) system (N) into symbols and strings (theorems) in a formal (either logical or mathematical) system (F), and then to decode the modeling results from the formal system into words meaningful to the observables in the real world system. Casti (1997) further argues that the key to understanding this process of formalization is to recognize that all notions of meaning (semantics) reside in the real world system N. In contrast, F consists of mere abstract symbols and the rules (syntax) for how these symbols can be manipulated to form new strings. The meaning of these symbols are extracted by decoding the strings back into N. The semantics of N is often rendered in induction and causation whereas the syntax of system F favors deduction and inferences. The goal of any modeling exercises is to find the most essential characterizations of system N first, and then to search for the most truthful representation of these characterizations in system F. Modeling is not successful if we fail to interpret the meaning of system F in the context of system N.

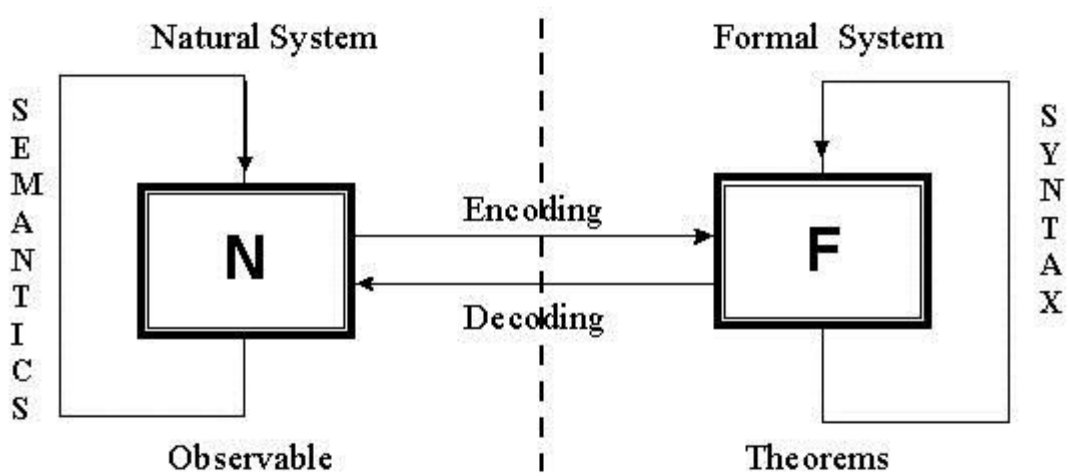


Figure 1. The Essence of Modeling [Modified after Casti (1994)]

2.2 Surprise-generating mechanism:

Surprise occurs when the results of F do not match those observables in system N. Surprise is the gap between our assumptions and expectations about the world and the way those events actually turn out. In essence, surprises are the end result of predications that fail. In an attempt to answer the challenging question as to why models fail, Casti summarized five main reasons for surprises by synthesizing the latest development in a vast array of disciplines such as quantum physics, computer science, biology, and mathematics. Although they are not mutually exclusive, the following five reasons are what Casti called the surprise-generating mechanism in complex systems:

Unpredictability:We are living in a world of essentially inconsistent phenomena, and long term prediction is impossible for complex systems. Chance must be treated as an actual cause for many things occurring in the real world..

Instability(the butterfly effect):Small changes in a system may cause large and catastrophic effects. These small changes are also implied throughout the system.

Uncomputability: Certain system behaviors defy explanations by rules. There is no prior reason to believe that any of the processes of nature and humans are necessarily rule-based. We could never see these processes manifest themselves in these surrogate worlds.

Irreucitbility: System behaviors cannot be understood by decomposing it into parts. Reductionism and atomistic view will lead to further illusion about reality. We must understand real world system as an organic whole.

*Emergence (Co-evolution):*Interactions among system components generate unexpected global system properties not present in any of the subsystems taken individually. Microlevel interactions between individual agents and global aggregate level patterns and behaviors mutually reinforce each other. Self- organizing patterns must be treated as both structure and process.

By combining a large amount of new discoveries from numerous scientific frontiers, Casti (1991; 1997) presented convincing evidence to support these five pervasive characteristics exhibited in both human and physical systems. Casti (1997) further argues that the science of surprise is how to deal with these five surprise-generating mechanisms. Geographers have also reported empirical evidences that are consistent with these five surprise-generating mechanism in both human and environmental systems (Dentrinos, 1990; Nijkamp and Reggiani, 1992; Phillips, 1993; 1995).

3. Evolving Versions of Urban Modeling

Casti's idea on modeling can be used as an organizing framework to examine the evolving versions of urban modeling. Comprehensive reviews on previous urban modeling efforts have already been made by Harris (1985), Batty (1994), and Wegner (1994). My intention here is to highlight the fundamental shifts in urban modeling during the past 40 years. I would like to use 1973 (the publication of Douglas Lee's article in JAPA) as a watershed year to group the highly diversified urban models into two versions (Table 1):

The first generation of urban models (1957-1973) can broadly be called the Lowry modeling tradition. The semantics rely on the implicit assumption of cities as simple systems which usually involve a finite number of individual elements with relatively weak interactions between them. The entities in the models are aggregated to predefined spatial units. The syntax of the first generation of urban models is based upon traditional linear, deterministic mathematical/statistical techniques such as those manifested in spatial interaction modeling, econometric methods, and optimization techniques borrowed from operations research (OR). The Lowry modeling heritage did not die after 1973 and modeling work following the Lowry tradition continues today all over the world (Wegner, 1994). During the past ten years, various modified versions of the Lowry type of models have been revitalized through their integration with GIS (Landis, 1995), but the core concepts and theories were developed during the 1950s and the 1960s. I have argued elsewhere that without critically examining the assumptions and theories, the integration of traditional urban modeling with GIS may be problematic just as putting old wine in new bottles does not make the wine any better (Sui, 1996; 1997a).

Table 1. Evolving Versions of Urban Modeling

Versions	
The first generation (1957-1973)	Cities were conceptualized aggregate; static; emphasi
The second generation (1973-present)	Cities are conceptualized a highly disaggregate ; dynar

The second generation of urban models (1973-present) tries to break away from the Lowry modeling tradition, with their emphasis on syntax (innovative mathematical concepts) and less concern on the semantics (new urban theories and urban realities). Different from the first generation of urban models, the second generation models conceive cities as complex systems which involve a large but finite number of intelligent and adaptive agents. The behaviors of these agents are contingent on the availability of information and subject to modify their rules of action based upon new information. This continual dynamism in the change of the behavior of agents makes the prediction and measurement by the old rules of science impossible. Therefore, the syntax of the second generation of urban models is characterized by the new concepts and theories in the non-linear dynamics. Beginning with Allen's work which introduced self-organizing and dissapitive structure theory into urban modeling in the late 1970s (Allen and Sanglier, 1981), urban modellers have jumped onto almost every major mathematical bandwagon invented after WWII, as evidenced by the introduction of catastrophe/bifurcation theory (Wilson, 1981), non-linear dynamics (Crosby, 1983; Bertuglia et al., 1990), fuzzy logic (Leung, 1988), Q-analysis (Gould, 1980), neural computing (Gimblett et al., 1994), chaos theory (Cartwright, 1991; Dendrinos, 1992), fractals (Batty and Longley, 1994), and cellular automata (Itami, 1994) etc. Several theoretical physicists have also contributed to this growth in urban modeling based upon non-linear dynamics and chaos theory (Maskse et al., 1995; Nagel and Barrett, 1997). In fact, TRANSIMS developed by physicists at Los Almos is perhaps the most ambitious urban model ever built. Unlike the first version of urban models, most of the second generation urban models are confined to the academia. Few models of the second generation of the modeling are operationalized for practical applications in decision making or policy impact analysis.

Although we can roughly identify these two different versions of urban modeling, Imust stress that the current urban modeling practice is characterized by a plethora of models in both versions. However, the philosophical shifts espoused by the second version of urban models are quite distinct and profound. I can detect at least the following two: 1). From a predominantly mechanistic view of cities based upon Newtonian physics to an organic view of cities based upon analogies in biology; 2). From a top-down approach to a bottom-up approach. The new approach emphasizes that cities are formed from more local actions without centralized planning or macro control. This may reflect a devolutionary trend in politics and a shift in planning ideology from instrumental rationality to communicative rationality.

4. Visions for Future Urban Modeling

Casti's framework not only enables us to gain a clear understanding of the evolving versions of urban modeling so far, but also, perhaps more importantly, stimulates our thinking on visions for future modeling efforts. I believe that our future modeling efforts should 1) focus on the new urban reality and develop new urban models and theories (semantics); 2) incorporate the paradigm of the non-linear science (syntax); and 3) embrace the latest computing paradigm for the efficient and effective implementation of urban modeling.

4.1 The new semantics of urban modeling

The first important element of the new urban modeling semantics is to recognize that we are living in fundamentally different kinds of cities than thirty years ago due to rapid technological advances after WWII. Accompanying each major revolution in transportation and communication technologies, American cities during the past two hundred years have been progressively transformed from the mercantile city (primarily influenced by wagon and sail technology in rivers and canals), to the early industrial city (relying on railways and sea-going vessels), and to the mature industrial city (dominated by automobile and air travel and long-distance communication). Since 1970, the post-industrial city - informational city resulting from the on-going revolution in telecommunications, computer, and media technologies - has emerged (Castells, 1989; Sui, 1997a). Cities have evolved from a mercantile city to a metropolis, to a megalopolis, and to a gigalopolis. Cities are experiencing not only a territorial expansion over geographic space, but also an increasing interaction and integration over cyberspace.

Although the urban forms and processes in information cities are still poorly understood, Graham and Marvin (1996) argued that the roles of space and time in urban life have been fundamentally altered as cities are being transformed from industrial cities to informational cities. Industrial cities tend to be spatially compact. Their goal is to overcome time with space, i.e. developed to make communications easier by minimizing space constraints to overcome time constraints. Whereas in the informational cities, telematic technology has completely destroyed the geocode key. Informational cities tend to be spatially diffuse. Their goal is to overcome space with time, i.e. developed to make communications easier by minimizing time constraints to overcome space constraints. The dramatic transformation of cities calls for a redefinition of the concept of a city. Inspired by Thrift and Olds (1996), I believe that cities nowadays can be conceptualized at least by the following four ways, each of which corresponds to a different physical and biological metaphor (Table 2). So far urban land use/transportation modeling has concentrated predominantly on the first two conceptions of cities. As cities are becoming more informational and further integrated with electronic spaces, we should give the last two conceptions of cities a higher priority. The first two conceptions of cities may be sufficient for industrial cities, but to completely understand information cities, we must combine all the four metaphors.

Table 2. Alternative Conceptualizations of Cities
[Modified after Thrift and Olds (1996)

Conceptualization of Cities as	**Biological Metaphors**	**Physical Metaphors**
Bounded Regions	Parts of Body	Physical Objects
Networks	Blood Veins	Roads/Highways
Space of Flows	Neural Networks	Energy Flows
Quantum States	DNA	Quantum Physics

This new urban reality poses many new challenges for urban land use modeling. I would like to mention only two here.

1). Telecommuting is one of the major trends in the U.S. now. According to a national survey, more 32 million Americans are working or running their businesses from home because of the increasing use of faxes, the Internet, cellular phones, etc. Obviously, what this entails is that residential land uses are increasingly blurred with commercial land uses. This means that we need to develop a new urban land use classification scheme for the information age rather than following the Anderson scheme. Telecommuting affects not only the traffic flows, but also the land use patterns in other parts of a city. My own study has shown that the increasing number of white- and pink collar workers working at home may be one of the main reasons for the high vacancy rate of downtown corporate office towers. In some cities, the vacancy rate of down town office towers is over 60%, and as a result, quite a few high-rise office buildings have been rented out for miscellaneous non-commercial purposes. It would be misleading to continue classify them as commercial land uses.

2). Traditionally, mapping the communication infrastructure and computer networks has not been treated very seriously in land use planning and monitoring. Even the new edition of Chapin's urban land use planning bible (Kaiser et al., 1995) makes no mention of it. But for information cities, computer and information infrastructures are crucial because of the changing nature of our cities. How cities are wired will be a very important factor influencing urban land use patterns. Because of the invisibility of information infrastructures and proprietary commercial interests, it is going to be a very challenging task to map the information infrastructures and make them an integral part of new urban models.

Obviously, we need a new semantics for urban land use modeling. This new semantic framework should unify land use structures (urban forms), land use functions (urban processes), and land use dynamics (urban policies to guide changes in forms and functions). This will require us to conduct thorough research on informational cities and examine how the current telematic revolution will manifest itself on the land. Without grounding modeling efforts solidly in the new urban reality, our sophisticated techniques may have little meaningful to say about the critical issues facing today's cities.

4.2. The new syntax for urban modeling

The new semantics of urban modeling demands that we must have a new syntax to model the complexity of the information city. Although linear and deterministic techniques are still applicable in certain situations, we need to expand our efforts to develop a coherent syntax for urban modeling using concepts and theories in non-linear dynamics. Incorporating insights gained from non-linear dynamics is the best way to handle the surprises in our future modeling efforts. Table 1 summarizes major techniques to handle each of the five surprise-generating mechanism. Although incomplete and overlapping among the five possible solutions, these solutions listed in Table 3 are a good starting point for the development of a unified framework to integrate those fragmented urban modeling works based upon non-linear dynamics.

Table 3. Possible Methods and Techniques to Handle Surprises in modeling

Surprise-generating Mechanisms	**Methods/Techniques**

Instability	Catastrophe/Bifurcation Theory
Unpredictability	Non-linear Dynamics/Chaos Theory
Irreducibility	Holistic Approach, Q-analysis
Uuncomputablity	Neural computing/Genetic Algorithms
Emergence	Self-organizing, Cellular Automata

I believe that chaos theory will play a central role in the new syntax for urban modeling. Chaos theory offers us a possibility of elegantly reconciling the simultaneous presence of complexity/irregularity and simplicity/regularity in a complex system. Chaos theory implies both apparent randomness out of order and order out of randomness. According to chaos theory, complex non-linear systems are inherently unpredictable, no matter how sophisticate or detailed the model may be. However, it is generally quite possible, even easier, to model the overall behavior of system. The way to express such an unpredictable system lies not in exact equations, but in representations of the behavior of the system -- in plots of strange attractors or in fractals. Pioneering works have already revealed that urban forms are essentially fractals in nature, and urban processes can be simulated as self-organizing cellular automata and neural networks (Batty and Longley, 1994; Clarke, 1997; Itami, 1994). We can expect that new developments in non-linear dynamics will play an increasingly important role as we switch our modeling focus from industrial cities as bounded regions and networks to information cities as space of flows and quantum states. In parallel to the unified semantic framework, we also need to develop a coherent syntactic framework integrating the concepts and theories listed in Table 3, just as what Alan Wilson did 30 years ago using the entropy maximization concept.

4.3 The new computing paradigm

The computational implementation of the new vision for urban modeling is inescapably tied to GIS. Rather than the stand-alone, layer-based approach, the emerging network-oriented feature-based GIS (FBGIS) through distributed computing and new protocols may represent the most ideal computing platform for the implementation of urban land use models.

Unlike the layer-based GIS in which we try to fit a map layer containing geographic entities into a Cartesian coordinate system (an absolute conceptualization of space and time), the FBGIS lends us a new conceptual framework to implement those alternative views of space and time and various new models depicting the physical and socio-economic processes in the real world (Tang et al. 1996). In a feature-based GIS, space, time and themes are defined as integral parts of a geographic feature instead of referencing all the entities into an arbitrary Cartesian grid. By providing direct access to spatial, temporal and thematic attributes, the FBGIS is not constrained to map and layered representations of geography and thus supports multiple dimensions of spatial/temporal events.

The other very important computing trend is to cultivate the interoperability of software products across distributed computing platforms (DCPs) according to the concept of the Open Geo-data Interoperatbility Specification (OGIS) (McKee 1996). The concept of OGIS has already stimulated new software development trends in the industry. Instead of developing a fully integrated GIS, more and more software vendors are engaging in developing a much leaner core module with numerous task specific, embeddable modules. These object-oriented, embeddable modules can not only be easily loaded into a core GIS package, but also can be seamlessly integrated with other application programs. In addition, with the explosive growth of both the Internet and the Intranet, the development of web-based software tools is necessary so that whoever has access to the Internet can run the program regardless of user's physical location. To implement new urban models using some of the web-based software development tools such as Java is definitely an area worth pursuing in the future. Those web-based modeling tools for urban development will enable citizen to more actively participate in the policy decision-making process. The development LUCAS is an exciting beginning for DCP-based land use modeling (Berry et al., 1996).

5. Concluding Remarks

The prospects for urban modeling are obviously more than just another twist in the change of spatial or temporal scales or a jump onto mathematicians' new bandwagon. New urban reality and new theories in science demand us to embark a fundamental paradigm shift for urban modeling at both semantic and syntactic levels. At the semantic level, we must realize that we are dealing with fundamentally different kinds of cities - informational cities as a result of the telematic revolution. Many old concepts and theories we are accustomed to are no longer applicable and new theories to this new urban reality have yet to be developed. At the syntactic level, the new development of science and technology during the later half of 20th century has provided us with a new set of language to describe and model various facets of urban reality. These new theories and concepts, as reflected in chaos theory, cellular automata, fractal geography, self-organizing theory etc., are rapidly coalescing into a non-linear science that challenges our deterministic, linear thinking that has existed since the time of Newton. Insights gained from preliminary studies using these concepts have enabled us to better understand the complexity of cities and the dynamics of land use patterns.

It would be entirely impossible to meet the dual challenges of urban modeling at both the semantic and syntactic levels without computers. Indeed, computers not only provide us with the tools to understand urban reality, but also are becoming an integral part of our cities we are trying to model using the same computers. To what extent we can succeed in this endeavor is a profound issue for all of us to ponder on for the days to come. We may never be able to eliminate surprises from our models, yet we can still hold out the possibility of creating something approximating what Casti called a science of the surprising.

References Cited

Allen P M, Sanglier M, 1981, "Urban evolution, self-organization, and decision making"*Environment and Planning A***13**167-183

Batty M, Longley P 1994*Fractal Cities*(London: Academic Press)

Batty M, 1994, "A chronicle of scientific planning: The Anglo-American modeling experience"*Journal of the American Planning Association***60**7-16.

Berry M W, Flamm R O,. Hazen B C, and MacIntyre R L, 1996, "Lucas: A System for Modeling Land-Use Change"*IEEE Computational Science and Engineering***3**24-35.

Bertuglia C S, Leonardi G, and Wilson AG (editors) 1990*Urban Dynamics*(London: Routledge).

Cartwright T J, 1991, "Planning and chaos theory"*Journal of the American Planning Association***57**44-56.

Castells M, 1989*The Informational City*(Oxford: Blackwell).

Casti J L, 1991*Searching for Certainty: What scientists can know about the future*(New York: William Morrow)

Casti J L, 1994*Complexification: Explaining a paradoxical world through the science of surprise*(New York: HarperCollins).

Casti J L, 1997*Would-Be Worlds: How simulation is changing the frontiers of science*(New York: John Wiley & Sons Inc.)

Clarke K C, 1997, "Long term urban growth prediction using a cellular automation model and GIS: Applications in San Francisco & Washington/Baltimore"*Paper presented at the International Workshop on GIS in Spatial Population Analysis & Regional Economic Development*, March 24-25, Hong Kong.

Crawford-Tilley J S, Acevedo W, Foresman T, and Prince Q, 1996, "Developing a temporal database of urban development for the Baltimore/Washington region"*Proceedings of ACSM/ASPRS Annual Meeting*, Washington D.C.,**3**101-110

Crosby R W (ed.), 1993*Cities and Regions as Nonlinear Decision Systems*(Boulder, CO: Westview Press).

Dendrinos D S 1990*Chaos and Socio-Spatial Dynamics*(New York : Springer-Verlag).

Dendrinos D S 1992*The Dynamics of Cities : Ecological determinism, dualism and chaos*(New York : Routledge).

Gimblett R H, and Ball, G L, and Guisse, A W, 1994, "Autonomous rule generation and assessment for complex spatial modeling"*Landscape and Urban Planning***30**13-16.

Gould P G, 1980, "Q-analysis, or a language of structure: An introduction for social scientists, geographers and planners,"*International Journal of Man-Machine Studies***12**169-199

Graham S, Marvin S, 1996*Telecommunications and the City: Electronic spaces, urban places*(London: Routledge).

Harris B, 1985, "Urban simulation models in regional science"*Journal of Regional science***25**:545-568

Itami R M, 1994, "Simulating spatial dynamics: Cellular automata theory"*Landscape and Urban Planning"***30**27-47.

Kaiser E J, Godschalk D R, and Chapin F S, Jr., 1995*Urban Land Use Planning*(4th edition) (Urbana: University of Illinois Press)

Kirtland D, Gaydos L, Clarke K, De Cola L, Acevedo W, and Bell C 1994 "An analysis of transformations in the San Francisco Bay/Sacramento area"*World Resource Review***6**206-217.

Landis J D, 1995, "Imagining land use futures: Applying the California urban futures model"*Journal of the American Planning Association***61**438-457

Lee D B, 1973, "Requiem for large-scale models"*Journal of the American Institute of Planners***39**163-178.

Leung Y, 1988*Spatial Analysis and Planning under Imprecision*(Amsterdam: North Holland)

Makse H A, Havlin S, and Stanley H E, 1995, "Modeling urban growth patterns"*Nature***377**608-612

McKee L, 1996, "OGIS spans distributed computing platforms"*GIS World***9**56.

Nagel K, Barrett C L, 1997, "Using microsimulation feedback for trip adaptation for realistic traffic in Dallas"*International Journal of Modern Physics*(forthcoming), http://studguppy.tsasa.lanl.gov/research_team/papers/ [Website for TRANSIMS]

Nijkamp P, Reggiani A, 1992 "*Interaction, Evolution, and Chaos in Space*" (Berlin: Springer-Verlag).

Phillips J D, 1993, "Spatial domain chaos in landscapes"*Geographical Analysis***25**101- 117.

Phillips J D, 1995, "Self-organization and landscape evolution"*Progress in Physical Geography***19**309-321

Sui D Z, 1996, "Urban forms, urban processes, and urban policies: A research agenda for the metropolis in the 21st century" In*Spatial Technologies, Geographic Information, and the City*, compiled by H. Couclelis (Santa Barbara, CA.: NCGIA Technical Report 96-10), 210-213.

Sui D Z, 1997a, "Reconstructing urban reality: From GIS to electropolis"*Urban Geography***18**74-89.

Sui D Z, 1997b, "GIS-based urban modeling: Practice, problems, and prospects"*Paper presented at the International Workshop on GIS in Spatial Population Analysis & Regional Economic Development*, March 24-25, Hong Kong (available from the author).

Tang A Y, Adams T M, and Usery E L, 1996, "A spatial data model design for feature-based geographic information systems"*International Journal of Geographic Information Systems***10**643-659.

Thrift N., Olds K., 1996, "Reconfiguring the economic in economic geography.*Progress in Human Geography***20**311-337.

Wegener M, 1994, "Operational urban models: State of the art.*Journal of the American Planning Association***60**17-29.

Wilson A G, 1981*Catastrophe Theory and Bifurcation*(Croom Helm, Beckenham, Kent)

A Land Transformation Model: Conceptual Elements, Spatial Object Class Hierarchies, GIS Command Syntax and an Application for Michigan's Saginaw Bay Watershed

Bryan C. Pijanowski[1]

David T. Long[2]

Stuart H. Gage[3]

William E. Cooper[4]

1 College of Natural Science
Michigan State University

2 Geochemical and Isotope
Department of Geological Sciences
Michigan State University

3 Spatial Analysis Laboratory
Department of Entomology
Michigan State University

4 Institute for Environmental Toxicology and
Department of Zoology
Michigan State University

1. Introduction

A suite of complex factors, including policy, population change, culture, economics, and environmental characteristics, drive land use change. Land use change is one of the most critical dynamic elements of ecosystems (e.g. Baker 1989, Richards 1992, Houghton 1994, Riebsame et al., 1994, Bockstael et al., 1995). Human-induced changes to the land often result in changes to patterns and processes in ecosystems such as alterations to the hydrogeochemistry (Flintrop et al. 1996), vegetation cover (e.g., Ojima et al., 1991), species diversity (Costanza et al., 1993) and changes to the economies of a community. It is for these reasons that issues surrounding land use are central to the concerns of local and regional resource managers and community land use planners.

Information about current land use patterns, the causes of land use change and the subsequent effects of these changes can be effectively communicated to resource managers, community planners and policy analysts using geographic information systems, predictive models and decision support systems (Cheng et al. 1996, Watson and Wadsworth 1996). The advancements in many geographic information system applications such as ARC/INFO (ESRI 1996), and the increased accessibility of spatial databases, makes developing simulation models within geographic information systems more feasible than even a few years ago.

This paper presents an overview of the modeling framework, systems approach, spatial class hierarchies, and a summary of the ARC/INFO commands used to construct our Land Transformation Model. A pilot Land Transformation Model has been developed that integrates a variety of land use change driving variables, such as population growth, agricultural sustainability, transportation, and farmland preservation policies for Michigan's Saginaw Bay Watershed. The pilot LTM utilizes a set of spatial interaction rules, which are organized into an object class hierarchy. The model is coded within a geographic information system with graphical user interfaces that allow users to change model parameters. Output of the LTM includes a time series of projected land uses in the watershed at user specified time steps.

2. ProjectObjectives

The objectives of the Land Transformation Project are to:

1. develop a spatial-temporal model that characterizes land use change in large regions;
2. create a model that is transferable in scope to other regions undergoing land transformation;
3. incorporate policy, socioeconomics and environmental factors driving land use change;
4. develop a pilot LTM that demonstrates proof of concept and that can be used to generate spatial and temporal aspects that can be generalized for the development of new model components;
5. apply a systems approach to model development; and,
6. use the model to test what-if policy scenarios.

3. Conceptual Elements

The Land Transformation Model (Pijanowski et al., 1995; 1996; in review) describes the influence of land use change on ecosystem integrity and economic sustainability of large regions. Conceptually, the Land Transformation Model contains six interacting modules (Figure 1): 1) Policy Framework; 2) Driving Variables; 3) Land Transformation; 4) Intensity of Use; 5) Processes and Distributions; and 6) Assessment Endpoints. All modules and submodules within the conceptual diagram are recognized not to be mutually exclusive; we use this diagram to illustrate main points and provide a foundation for the description of more detailed model components. The pilot LTM that is described below contains two of the six LTM modules, driving variables and land transformation. The spatial extent of the LTM can be any definable region; however, because future model developments will be focused on coupling land use change and hydrogeologic and geochemical processes, we give precedence to watersheds as the spatial extent in LTM applications.

The *Policy Framework* module of the LTM organizes the goals for the watershed's stakeholders who include resource managers, private and corporate landowners, and local land use planners.

Stakeholder goals may include: control of pollutant inputs, ecological restoration, habitat preservation, improving biodiversity and biological integrity, and facilitating economic growth. Within this framework, many stakeholder goals are under certain types of constraints (e.g., economic, environmental), are made with certain expectations of outcomes and with specific spatial and temporal scales in mind. For example, a township land use planner is likely to be making decisions within his/her own township. Likewise, a state or federal government resource manager might be concerned about areas that encompass several counties.

The Land Transformation Model (LTM) contains three general categories (Figure 1) of *Driving Variables*: Management Authority, Socioeconomics and Environmental. Management Authority includes the institutional components and policies of land use. Land ownership is an important component in this module of the model since state and federally-owned lands (e.g., state and federal forests, parks and preserves) need to be excluded from development. Socioeconomic driving variables include population change, economics of land ownership, transportation, agricultural economics and locations of employment. Environmental driving variables of land transformation are: (1) abiotic, such as the distribution of soil types and elevation; and, (2) biotic, such as the locations of endangered and threatened species, or the attractiveness of certain types of vegetation patterns in the landscape for development. Driving variables may contain intercorrelated subcomponents; hence the model can be hierarchical. For example, the farming socioeconomic system in the Saginaw Bay Watershed application of the pilot model is composed of farm-size dependent economics, farmer demographics and environmental influences on farm productivity.

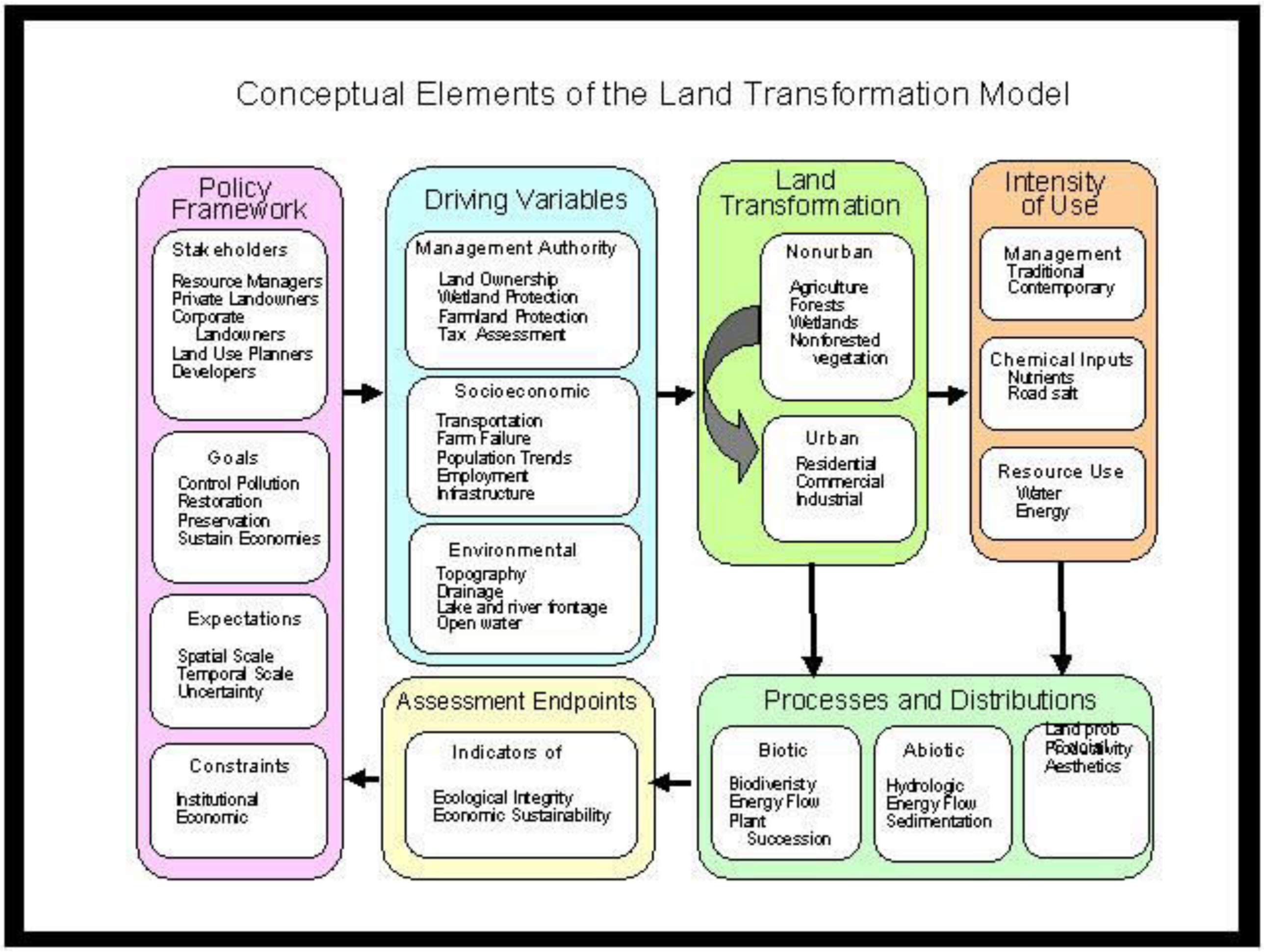
Land Transformation is characterized by change in land use and land cover. Land use describes the anthropogenic uses of land as its affects ecological processes and land value (Veldkamp and Fresco 1996). Land uses that we consider at the most general level are: urban, agriculture/pasture, forest, wetlands, open water, barren and non-forested vegetation. Land cover characterizes the plant cover of associated land use and is thus not mutually exclusive of land use. Land cover types that are considered include: types of agriculture (row crops versus non-row crops), deciduous and coniferous forests, and non-forested vegetation.

Within each land use, we consider *Intensity Of Use*such as land management practices, resource use and human activities. Intensity of use can be measured as chemical inputs to the land to increase its productivity (e.g., herbicides), chemical inputs as it results from human activities (e.g., salting of roads), and natural resource use (e.g, subsurface water for irrigation, per unit area energy consumption and forest harvesting). Socioeconomics, policy and environmental factors will also drive the intensity of use as well.

Changes in land use and cover, and intensity of use, alter *Processes*(e.g., hydrogeologic and geochemical)*and Distributions*of plants and animals in ecosystems. Processes that we are interested in characterizing include groundwater and surface water flows, chemical and sediment transport across land and through rivers and streams, geochemical interactions and fluxes such as nutrients (nitrogen and phosphorus). Land use and land cover will affect the types and numbers of animals inhabiting areas.

*Assessment endpoints*are indicators of ecological integrity and economic sustainability. These assessment endpoints are used to quantify the nature of changes in landscapes. It is important that assessment endpoints be: 1) relatively easy to quantify, 2) unambiguous, 3) correlated with changes to land use; and, 4) reflect qualitative aspects of landscapes. These assessment endpoints provide input to the decision making process by watershed stakeholders.

Figure 1. Conceptual Elements of the Land Transformation Model

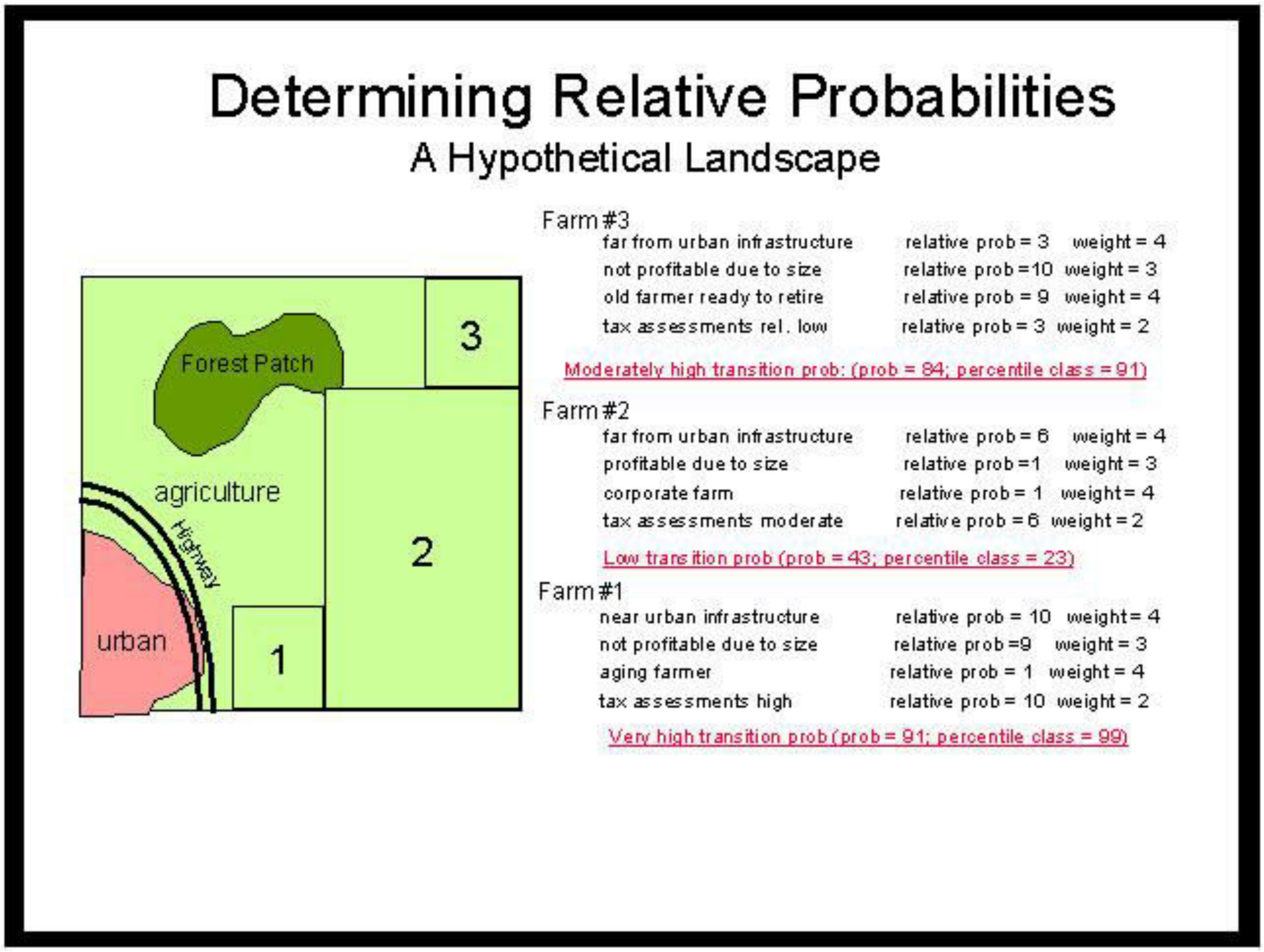


4. Spatial Framework

Land use and features (roads, rivers, etc.) in the watershed are characterized in the pilot LTM model as a grid of cells. Each cell is assigned an integer value based on land use (e.g., urban, agriculture, wetlands, forest) or land feature. Driving variable calculations produce land use conversion probabilities for each cell. The GIS is used to perform these driving variable calculations, integrate all driving variable conversion probabilities and produce future land use maps for the entire watershed.GIS calculations in grids commence at the upper left corner of the grid and end at the lower right corner of the grid. In the Saginaw Bay Watershed application of the pilot LTM, up to 5.2 x 10^7 cells are contained in each grid.

Figure 2 illustrates conceptually how land use transitions are determined in the Land Transformation Model. This hypothetical landscape contains three agricultural parcels: a small parcel near a highway, a large parcel some distance away from the highway and another small parcel a relatively large distance away from the highway. The drivers to land use change operate on these parcels differently depending upon the spatial relationships of the parcel and the drivers. For example, parcel #1 is under pressure for development due to its: proximity to a highway; proximity to urban infrastructure such as city water and sewers; proximity to high density employment centers found in the urban areas; and, due to its size, the farm is not likely to be profitable. Furthermore, its landowner may also be older and because few younger people are not entering agriculture, it is at a high risk of being converted out of agriculture and into an urban use. The second farm, as indicated by parcel # 2, is held in agriculture by the nature of its ownership (i.e., corporate). Parcel #3 in this figure has a higher probability of converting to urban land use because of the demographics of the owner, and the size of the parcel.

Figure 2. Relative Land Transition Probabilities



In the LTM, we use the GIS to make spatial calculations between drivers of land use change and cells being considered for land transition. The values resulting from these calculations are converted to*relative land transition probabilities.* Relative land transition probabilities that are used range from 1 (lowest probability of undergoing transformation to urban land use) to 10 (greatest chance of being converted to urban land use). Creating these relative probabilities from absolute GIS calculations requires: 1) spatial scaling or assigning relative transition probabilities based on absolute values; and, 2) making adjustments to state transition patterns. The types of spatial scaling and state transitions considered in the LTM are described below as part of the presentation of spatial class hierarchies.

In addition to calculating relative land transition values based on: 1) spatial interactions of drivers; and, 2) cells within a parcel, relative weights for each driving variable are assigned and these are then used to calculate urban transition values for each cell in an area. All land transition probabilities and weights for each driving variable are then integrated with the GIS for each location. Values are then placed into equal area percentile classes. Cells with the greatest percentile value are assumed to transition first to urban. The number of cells for each future transition is based on the per unit area requirements for urban given population growth projections for an area (township, county or entire region). The number of cells that meet the demands for each successive projection (e.g., decades) are then transitioned to urban. A more detailed description of the model calculation process can be found in Pijanowski et al. (in review).

5. Spatial Class Hierarchies

Figure 3 illustrates the LTM Spatial Object Class Hierarchy. There are six principal spatial classes in the LTM: interactions, resolution, spatial scaling, state transitions, landscape features and the number of subdrivers. Each of the principal spatial classes in turn are composed of several subclasses, which may be further divided into more refined spatial objects. The terminal positions of the space object classes become rules from which software modules are developed within the geographic information system.

5.1. Spatial Interactions

Spatial interactions used in the LTM are: neighborhood, distance, patch size, and site specific characteristics.*Neighborhood*spatial interactions are based on the premise that trends and patterns in neighboring locations influence a cell's land use transition probability. Neighborhood interactions can also vary in size, from those that only occur among proximal locations to large neighborhoods that encompass large areas (counties, subwatersheds or the entire watershed). We also recognize that the shape of neighborhood's may differ, from square, circular to irregular (e.g., watershed catchment).

*Distance*functions are the second type of spatial interactions used to characterize driving variables of land use change. We use the geographic information system to calculate the distance of locations in the watershed from landscape features (e.g., roads, rivers, employment centers) and convert these raw values into relative probabilities of land transformation (conversion rules are described under*state transitions*below).

*Patch Size*is based on the principle that the size of a parcel of land held by an owner has an influence on whether a land use conversion is eminent. For example, farm size in the United States impacts profitability such that small farms cannot compete with larger farms who can invest in advanced farm machinery, etc. Thus, small farms are at greater risk of failure and hence being converted to a non-agricultural use than larger farms.

Finally,*site specific characteristics*are also important to land use conversion. Certain characteristics (e.g. soil type or elevation) make each site suitable or unsuitable for a particular land use. Policy may also influence site specific characteristics of land transformation by either locking land in a specific land use or promoting its conversion.

5.2. Resolution

Examples of the*resolution*spatial object class in the LTM include those for cell size. Four different resolution classes are used in the LTM: parcel (30 x 30m), plat (100 x 100m), block (300m x 300m), and local (1km x 1km). These rules were developed to characterize certain processes such as land ownership changes which occur at relatively high resolutions (e.g., 30 x 30m) and hydrologic dynamics that occur at more coarse resolutions (e.g. 1 km x 1km resolution). Selection of resolution is also determined by the resolutions of databases available to study a process or pattern (e.g., land use is 30x30m as it might be developed from Landsat TM). We integrate multiple grids in the GIS using either the*minof*or*maxof*option in the*setcell*function in ARC/INFO GRID.

5.3. Spatial Scaling

Creating these relative probabilities from absolute GIS calculations requires: 1) spatial scaling or assigning relative transition probabilities based on absolute values; and, 2) making adjustments to state transition patterns.

Spatial scaling to convert all raw GIS calculations (e.g., distances) to relative probabilities is accomplished using the*slice*function in ARC/INFO GRID. Two options of this function are employed: equal area or equal class sizes. The former option of the*slice*function produces driving variable grids with equal numbers of cells with values between 1 and 10. The latter option provides driving variable grids with equal size classes between the largest and smallest values in the entire grid

Relative transition probabilities can be assigned based on absolute values rather than using spatial scaling routines as described in the previous paragraph. For example, in the Saginaw Bay Watershed application of the pilot LTM, relative transition probability values of 10 were assigned to all cells 30 m on either side of state and county roads within 100 m of highway intersections; all cells 30 m around county and state intersections were assigned values of 7; and all cells on either side of state and county roads were assigned values of 5.

5.4. State Transitions

Two different state transition adjustments made in the LTM. First, the direction of the relationships between the spatial scaling routine result and land transition probability may be positive or negative. For example, land closer to road intersections have the greatest probability of conversion to urban. The GIS is used to calculate the Euclidean distance of cells from the nearest road intersection and these values are then spatially scaled to create grids with relative probability values where the largest values are assigned 10s and the smallest values a 1. However, land closet to a driver such as a road have the greatest probability of conversion to urban; thus, there is a*negative*relationship between the result of the spatial scaling and the degree of urbanization. We invert these transition values using the following simple expression:

outgrid = 11 ingrid

where outgrid is the inverted driving variable grid and the ingrid is the input grid that contains values from 1-10.

The relationship between a spatial calculation and the influence of this result on urbanization can also be linear or nonlinear (Figure 3). In the case of nonlinear relationships, prior to spatial scaling, all values are adjusted by transforming values using the exp() function in ARC/INFO GRID. The equal size class option of*slice*is only used for spatial scaling of these state transitions.

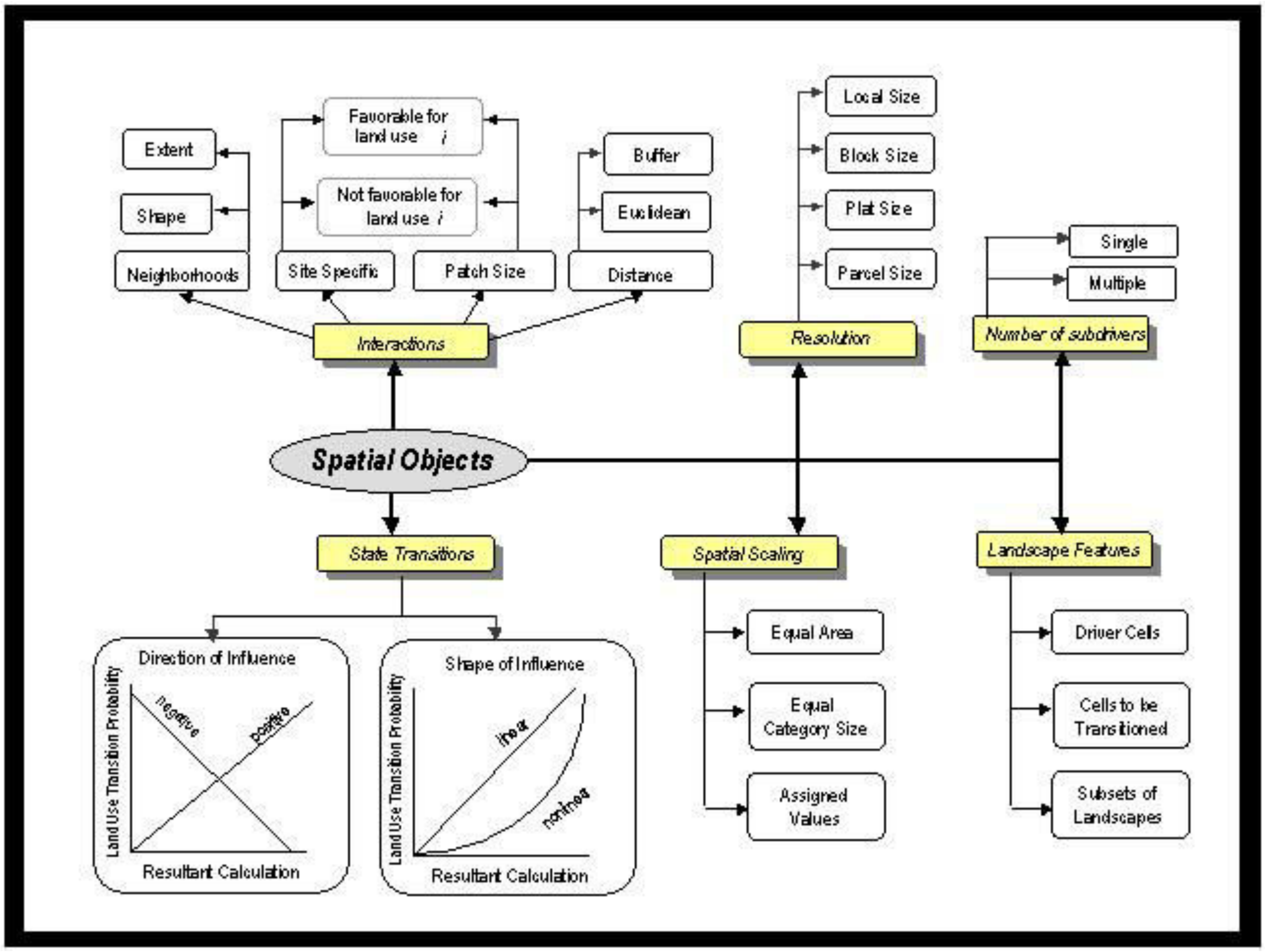
5.5 Landscape Features

The fifth type of spatial class objects in the LTM are landscape features. In many instances, the presence of absence of a feature in the landscape is important in the calculation of a land transition probability. For example, the relative density of farms in a local area are derived by producing a map of the presence (coded as 1) or the absence thereof (coded as 0) of agriculture in all locations in the watershed. Features are also cells that are considered for transition and those that are drivers of land use change.

5.6 Number of Subdrivers

Single or multiple layers are required to develop a driving variable. Multiple layer examples include those subdrivers that influence farm failure such as farm size, farmer age, amount of available surrounding arable land, soils, climate, and farm infrastructure (e.g., drains).

Figure 3. Spatial Objective Class hierarchy



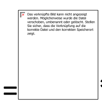
6.0. GIS Framework and Syntax

6.1. GIS Integration Schematic

Figure 4 illustrates how the GIS is used to produce land use projection maps. The first step is to create driving variable grids that contain values representing relative transition urban probabilities. This process may first require producing grids that contain information about the absence (cell value = 0) or presence (cell value = 1) of a feature (e.g., road) or land use type (e.g., agriculture); several grids may be integrated to produce the necessary input layer (Figure 4; Step 1A). Spatial calculations (e.g., neighborhoods, Euclidean distances) are performed (Figure 4; Step 1B) on the input grids so that resultant raw values (e.g., distance a cell is from a driver cell) are stored in each cell in the grid (Figure 4; Step 1C). These raw values are then scaled (Figure 4; Step 1D) so that there are an equal number of values between 10 (greatest probability on urbanization) and 1 (least probability for urbanization). This process produces driving variable grids (1E) that are then multiplied by a driving variable weight (1F). All driving variable grids are then summed (i.e., all cells for each location are added together) and this sum is stored in a final integrated driving variable grid (Figure 4; Step 2).

Cells within the grid that are identified as non-buildable due to policy (e.g., development rights have been restricted) or ownership (e.g., land is state or federally owned) are created (Figure 4; Step 3A) so that non-buildable cells are assigned value of 0 and potentially buildable areas assigned values of 1. All of these grids are integrated by multiplying them together so that a single building exclusion grid is produced Figure 4; Step 3B).

An urban pressure grid is produced as part of Step 4 in the GIS integration process; this is created by multiplying the building exclusion grid with the integrated driving variable grid. A nonurban grid (nonurban cell = 1; urban = 0) is used to multiply with the urban pressure grid. This step results in an area to be transformed grid (Figure 4; Step 5A) that contains integrated driving variable values for all nonurban areas. Values in the nonurban areas are then scaled (Figure 4; Step 5B) into percentile classes so that each percentile is represented by an equal number of cells (i.e., each value between 1 and 100 contains equal areas) in the grid labeled as 5C. The number of cells transformed to urban is determined by calculating a critical threshold value (Step 5D). Estimating the appropriate critical threshold value can be accomplished as follows. First, the amount of future urban land is determined using population growth projections and per capita urban land requirements:

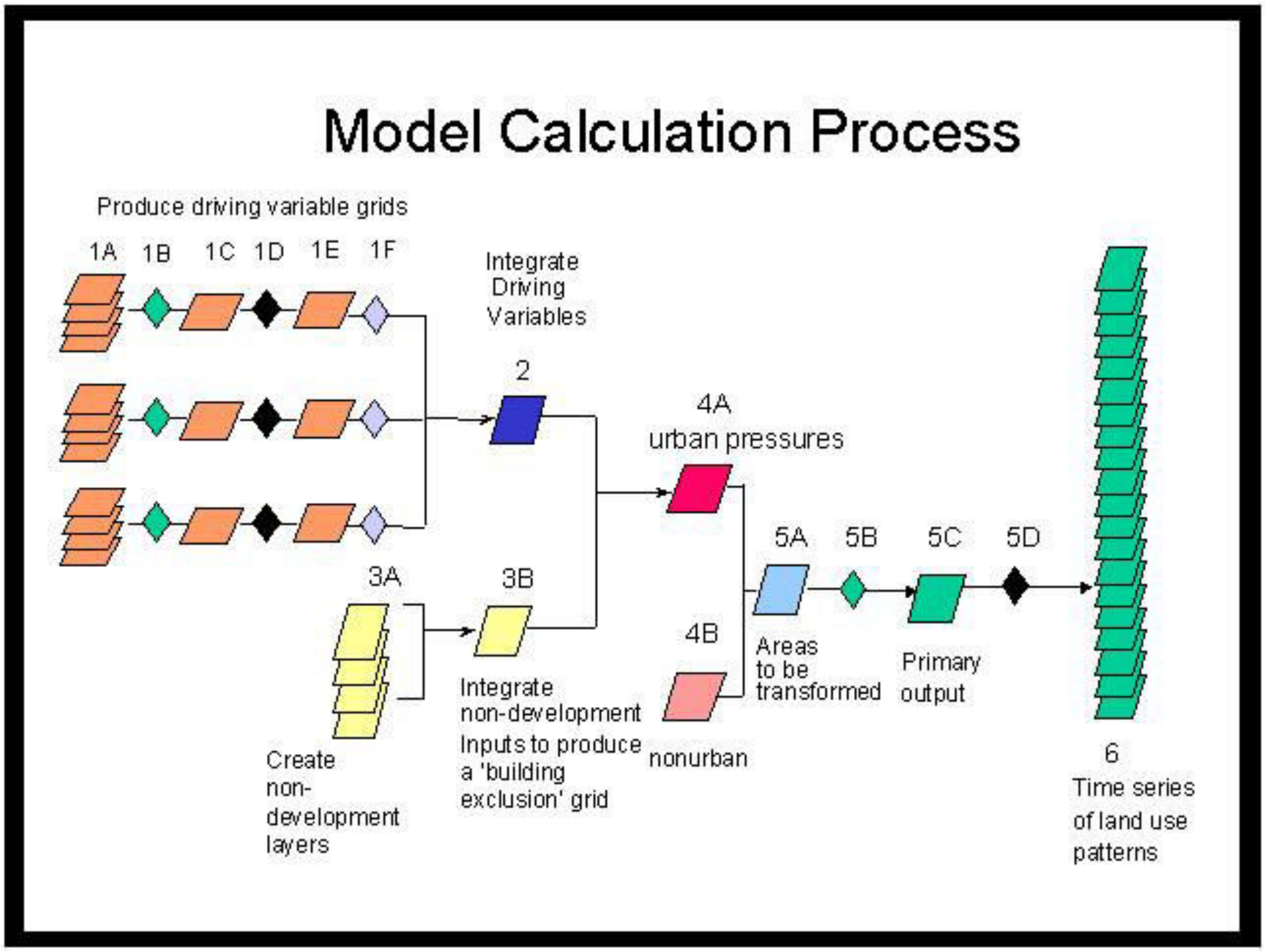
U (t) =  * A (t). (1)

where *U* is the amount of new urban land required in the time interval *t*, *P* is the number of new people in any given area in a given time interval and *A* is the per capita requirements for urban land. The critical threshold value is then simply a proportion of the current nonurban land use to the amount of new urban land use required in the future:

$$C(t) = 100 - \{[U(t) / N] * 100.0\} \quad (2)$$

where *N* is the amount of current non-urban land use that can be developed in the future, expressed as a percent. Note that *N* is also a function of non-buildable area. Future land use grids are produced that step through the critical threshold values (Step 6).

Figure 4. GIS Integration Schematic



6.2. ARC/INFO Commands

The cell-based spatial modeling software module found in ARC/INFO GRID (ESRI 1997) is used to model land use change. The GRID functions used most often in the LTM are:

setcell: this sets the cell size prior to any calculation;

focalsum: this calculates the sum of all values of any given area surrounding each cell and stores that value in that cell. This function allows the user to set the size and the shape of the neighborhood.

eucdistance: this calculates the distance between two cells in the landscape and stores that value in each cell;

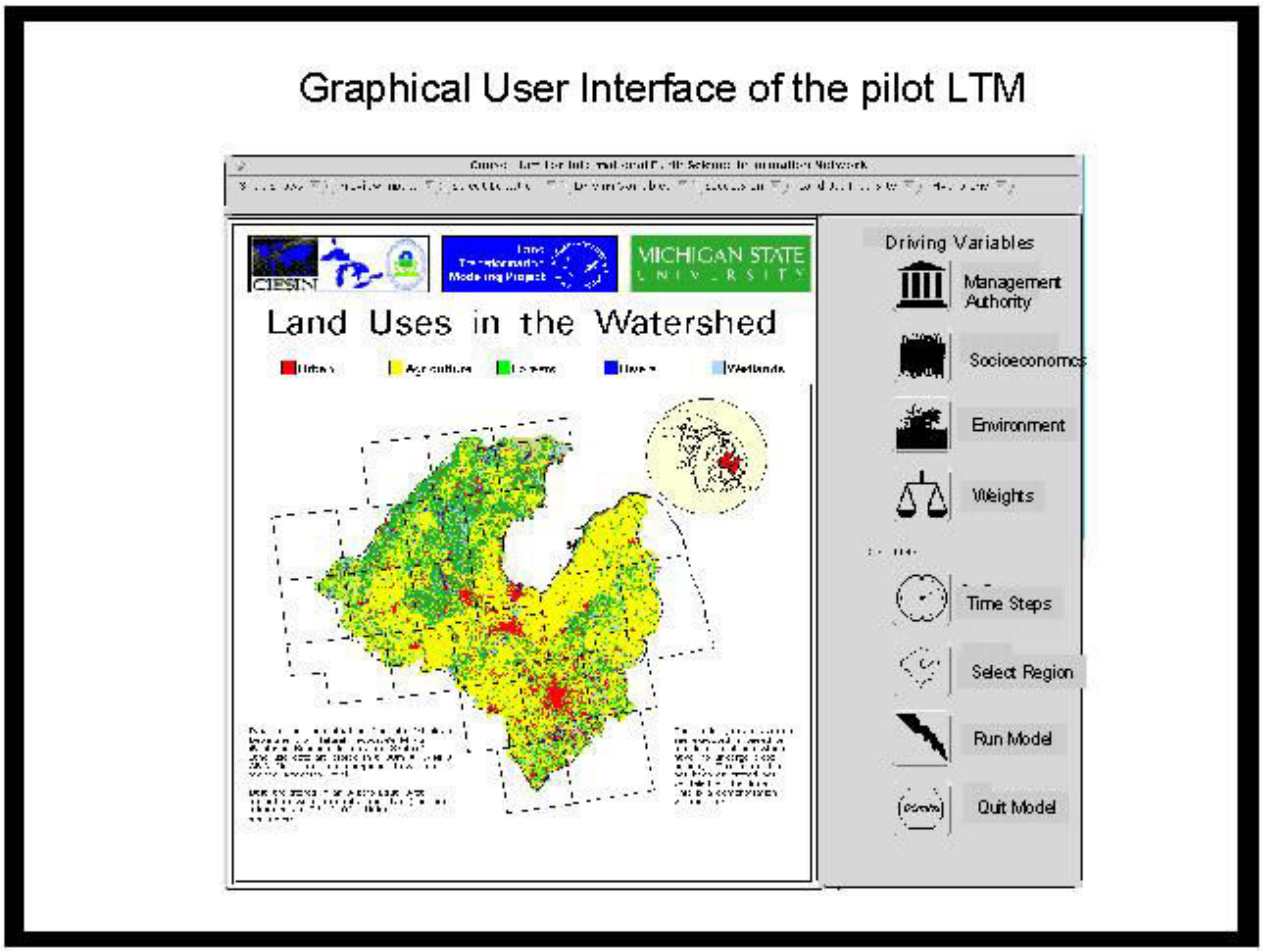
zonalarea: this calculates the area of contiguous cells which have the same assigned value;

test: this performs a Boolean operation on input data and stores a 1 if the given expression is true and 0 if it is false; and, *slice*: this changes all input values to an integer ranking based on the amount of area assigned to each ranking or the size of the category of values.

6.3. Model Interface

Figure 5 shows a sample user interface of the Land Transformation Model. This interface was developed using the Formedit GUI development tool in ARC/INFO in the OpenWindows UNIX environment. The interfaces allow users to set values for driving variable calculations (e.g., cell size, neighborhood extent) as well as provide access to visualization and output analysis tools.

Figure 5. Sample Graphical User Interface of the pilot LTM



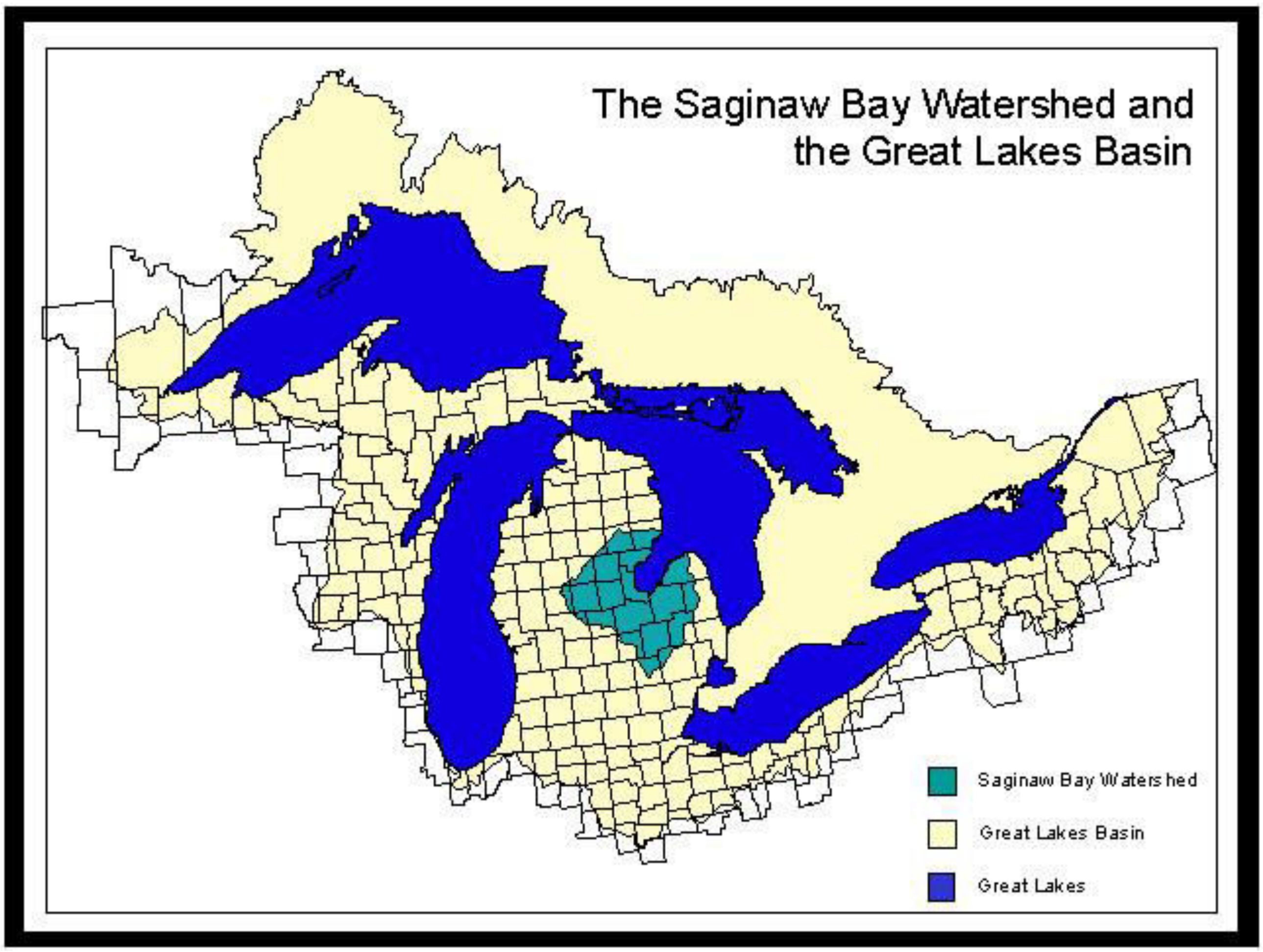
7.0 Model Application

7.1 Site Description

The Saginaw Bay Watershed (SBW) is one of the largest watersheds in the Great Lakes (Figure 6) area covering approximately 15,000 km^2(15% of the total area of the state of Michigan).

The SBW is composed of 10 smaller watersheds which are further divided into 69 subwatersheds. The principal river in the watershed is the Saginaw, which is only 47 km long; however, it drains 28 rivers and streams and nearly 73% of the watershed (MUCC 1993). There are three major tributaries of the Saginaw River: the Cass River to the east, the Flint River to the south, and the Titabawassee River to the west. The major cities within the watershed include Flint, Saginaw, Bay City, Midland, and Mt. Pleasant. There are 22 counties, 42 cities, 50 villages, and 277 townships in the watershed. Each municipality (e.g. township or cities) is given the authority to govern their own land use. Over 1.1 million people live in this watershed.

Figure 6. Saginaw Bay Watershed

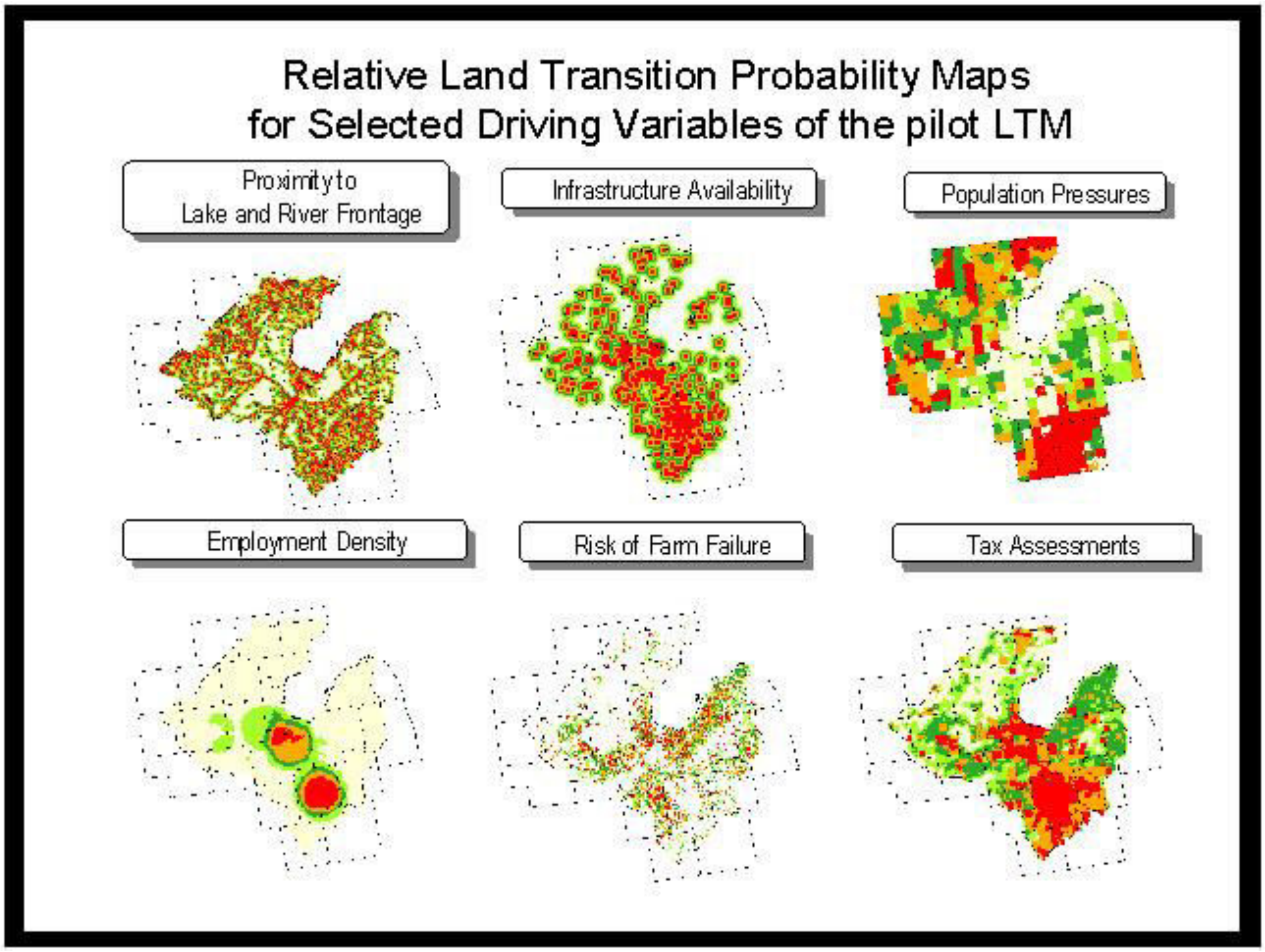


Agriculture is by far the most common land use in the SBW (46%), followed by forested areas (27%), and open vegetation (non-forested vegetation) (11%). In the SBW, fewer than 8% of the cropland is under conservation tillage compared to the statewide average of 40%. The lack of conservation tillage practices has created a situation of massive soil erosion due to wind and water. Urban use makes up only 6.6% of the entire area. Within urban areas, residential areas comprise 67% of the urban area. The other major urban uses are commercial (9%), transportation (8%), and industrial (4%). Topography does not vary considerably in the watershed. Areas near the mouth of the Saginaw Bay differ by less than 3 meters from 10 miles inland. As a result, flow of the major streams in the Saginaw is relatively slow; in some cases, the Saginaw River has been known to flow in the reverse direction during strong northeasterly winds.

7.2. Pilot LTM Driving Variables

We have used the LTM conceptual diagram (Figure 2) to develop a pilot GIS-based simulation model that forecasts land use in the Saginaw Bay Watershed using*policy*,*socioeconomic*and*environmental*driving variables. This model represents two of the six LTM modules. This pilot model's driving variables are: land ownership; the state's farmland preservation act and its affect on farm to urban conversion; the state's wetland protection act; the effect of the state's property tax assessment method on farm failure; the Suburban Control Act; local and regional population change; economics of land ownership; transportation effects on urbanization; local and farm level agricultural economics; location and density of employment opportunities and social factors that affect farm failure; the presence or absence of buildable soils; the affects of drainage system on agricultural performance; and, the relative attractiveness of several landscape features for urban development. Figure 7 illustrates some of the driving variable calculations results. A more detailed description of the driving variable calculation formulation can be found in Pijanowski et al. (in review).

Figure 7. Sample pilot LTM Driving Variables

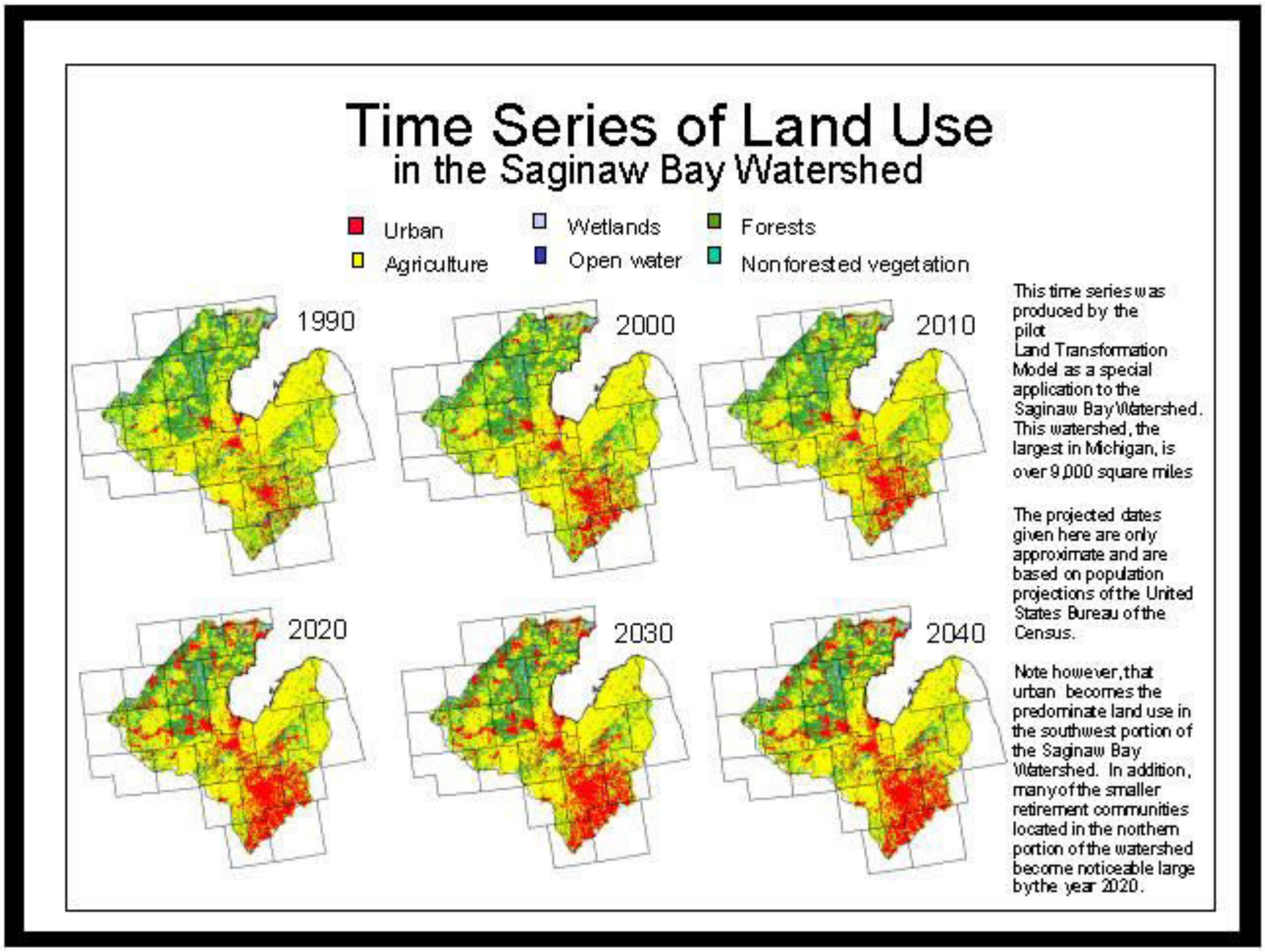


8.0 Results and Discussion

8.1 Summary of Results

The pilot LTM was executed without assigned weights to the 13 driving variables listed above (Figure 8). The critical threshold values for each 10 year time step was determined from State of Michigan population projections for the next 50 years. In addition, the base land use map that was used was developed by synthesizing land use polygons from 350 townships in the watershed from the Michigan Resource Information System (MiRIS). Land use from this database is current only to 1980. Thus, the first projection created a land use map for 1990. In the near future, the pilot LTM will be calibrated by conducted an historical forecast in order to attempt to predict current land use conditions. A current land use map for the entire watershed is planned to be completed by August of 1997.

Figure 8. Zoom of LTM Execution Results in Genessee County



9.0 Acknowledgements

We gratefully acknowledge funding from the Consortium for International Earth Science Information Network (CIESIN), University Center, Michigan, through Cooperative Agreement (CX821505) between the U.S. Environmental Protection Agency and CIESIN. William Cooper and Tom Edens, two co-PIs on the project, provided substantial input on the development of the model. We would also like to thank Bob Worrest, Mike Thomas, and Bob Bourdeau of CIESIN; Charles Bauer and James Bredin of the Michigan Department of Natural Resources for helpful comments during the model development phase. Dennis Tenwolde reviewed earlier versions of the manuscript. Amos Ziegler, Katie Jones, Tom Sampson, Gary Icopini, John Abbott and Mark Rousseau helped to prepare a variety of the spatial databases used in this project.

10.0 References

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

Clarke, K.C., Hoppen, S., Gaydos, L., .1996b. 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.

Costanza, R. 1989. Model goodness of fit: a multiple resolution procedure. Ecological Modeling 47: 199-215.

Costanza, R. and T. Maxwell. 1994. Resolution and predictability an approach to the scaling problem. Landscape Ecology 9(1):47-57.

Costanza, R. F. Sklar, M. White. 1990. Modeling coastal landscape dynamics. Bioscience 40 (2): 91-107.

Costanza, R., and F. Skalr. 1985. Articulation, accuracy and effectiveness of mathematical models: a review of freshwater wetland applications.

Environmental Systems Research Institute. 1996. Cell-based modeling with GRID. Environmental System Research Institute, Atlanta, Georgia.

Flintrop, C., B. Hohlmann, T. Jasper, C. Korte, O. Podlaha, S. Scheele, and J. Veizer. 1996. Anatomy of pollution: streams of North Rhine-Wesphalia, Germany. Am. J. of Science 296: 58-98.

Kolak, J.J., D.T. Long, J.M. Matty, G.J. Larson, and D.F. Sibley (1996) Groundwater, stream, lake dynamics: Saginaw Bay Watershed. International Association for Great Lakes Research Annual Meeting 1996.

Michigan Department of Natural Resources. 1994. Saginaw Bay Watershed Prioritization Process. MDNR. 65 pp.

Michigan United Conservation Clubs. 1993. Saginaw Bay Watershed Land Use and Zoning Study. 308 pp.

Pijanowski, B. C., T. Machemer, S. Gage, D. Long, W. Cooper and T. Edens. 1995. A land transformation model: integration of policy, socioeconomics and ecological succession to examine pollution patterns in watershed. Report to the Environmental Protection Agency, Research Triangle Park, North Carolina. 72 pp.

Pijanowski, B. C., T. Machemer, S. Gage, D. Long, W. Cooper and T. Edens. 1996. The use of a geographic information system to model land use change in the Saginaw Bay Watershed. Proceedings of the Third International Conference on GIS and Environmental Modeling, Sante Fe, New Mexico, January 21-26, 1996. GIS World Publishers on CD-ROM.

Pijanowski, B.C. 1997. Issues of space and time scales in changing landscapes: systems approach and the use of geographic information systems. In Lecture Series in Landscape Ecology. Larry Harris, ed.

Pijanowski, B. C., T. Machemer, S. Gage, and D. Long. In press. A land transformation model: conceptual and analytical framework and its application to Michigan's Saginaw Bay Watershed. To Ecological Modeling.

Sklar, F. And R. Costanza. 1991. The development of dynamic spatial models for landscape ecology: a review and prognosis.

The Development of a Land Cover Change Model for Southern Senegal

Eric C. Wood, John E. Lewis, G. Gray Tappan, Ronald W. Lietzow

USGS EROS Data Center / Hughes STX Corp

Sioux Falls, SD 57198

ABSTRACT

Temporal and spatial Markov models were developed for characterizing land use/land cover change in south-central Senegal, as an early stage in the development of models for lulc change forecasting. The modeling effort is a component of a project carried out by the EROS Data Center and funded by the U.S.Geological Survey and the Senegal mission of the U.S.Agency for International Development. The Markov model results were compared to results obtained using other spatial analysis techniques including join-count statistics and landscape metrics. Though landscape change appears not to be strictly Markovian, pure and modified versions of both the temporal and spatial models appear to have potential for simulation of lulc cover.

INTRODUCTION

The increasing rate of land use/land cover change is one the of the most important ecological issues in Africa (de Graf, 1993). Because the majority of African countries have rural economies and depend heavily upon their natural resources (e.g. for food, fuelwood, commodity exports), degradation of those resources can result in rapid declines in standards of living.

This has been particularly apparent in the Sahelian countries. As population pressures, international economies, and mechanized agriculture have become more prominent, Sahelian Africans have been forced into land use practices that are often inappropriate and result in degradation of soils and forests.

Landscape change also has an influence at regional and global scales. Deforestation, desertification, erosion, and many other types of land cover change contribute directly to a loss in biodiversity and very likely global climate change. Reduced resource bases can lead to regional food shortages, political instability, and the humanitarian concerns related to both.

Of particular importance are the types and degree of changes in the landscape, which must be determined before any causal relationship can be postulated. Long term monitoring of key indicators is a prerequisite to such a study, with attention paid to physical, biological, and social phenomena (Wasson, 1987). Once the type and extent of land cover change has been determined, modeling efforts can be initiated that describe it and forecast future change.

The following paper discusses the early stages of the development of such a model. Specifically it addresses the use of Markov processes in characterizing temporal and spatial changes in land use/land cover as part of a long term environmental monitoring project in the West African country of Senegal.

BACKGROUND

The EROS Data Center (EDC) of the U.S.Geological Survey and the Senegal mission of the U.S.Agency for International Development (USAID) have begun a pilot effort to develop a long term environmental monitoring system. A major goal of the project is to design, test, and institutionalize a framework for monitoring changes in Senegal's natural and agricultural resource base.

As part of that effort, EDC is working with the Government of Senegal's Ministry of the Environment and Protection of Nature to determine the extent of land use/land cover change; the biophysical, social, economic and political factors related to that change; and developing models that both describe existing changes and forecast future changes in land use/land cover. Two key requirements of such models are that they are technically straightforward enough to be applied in interested government agencies and yet accurate enough to be of use in providing realistic simulations for policy makers.

A prerequisite in modeling land use/land cover change is the ability to characterize existing change both temporally and spatially using empirical data. This step is necessary for forecasting future change. In order to develop such a characterization, Markov models were chosen. In the past such models have been used for urban land use/land cover change modeling (Bell, 1974; Bourne, 1976), forest and vegetation succession modeling (Horn, 1975; van Hulst, 1979; Hall, 1991; Usher, 1992), and more recently in modeling landscape change (Baker 1989, Turner 1990, Flam and Turner 1994, Boerner et.al., 1996).

Markov models have substantial scientific appeal. They are mathematically compact, easily developed from observed data and serve as an effective tool for simulation exercises. However, these models are not without their problems, indications of which will be discussed later.

Other studies have attempted to model land use/land cover change in the region. Perhaps the most well known effort in Senegal is by Lake (1979). In this study, Lake projected both forward and backward the extent of agricultural expansion from 1929 to 2004, based on land use/land cover data from 1954 to 1979. More recently a model was developed by Gilruth et.al. (1995) to simulate the dynamics of shifting cultivation in the Guinea Highlands (Futa Djallon). Both studies provide a useful backdrops for present efforts.

STUDY AREA

The overall study area consists of four departments located in south-central Senegal (Kaffrine, Tambacounda, Kolda and Velingara). There are 30 departments in all of Senegal, these four being of particular interest in that they have abundant natural resources, fall within the zone of viable rainfed agriculture, and as a result are undergoing significant land use/land cover change. The initial investigation was carried out in the department of Velingara (Fig. 1) and is the focus of this paper.

The landscape of Velingara is a mosaic of upland dry Sudanian woodland, riparian or moist gallery forest in the fossil river valleys, and clearings for agriculture, with no major urban centers. Most of the department, 66%, is in the Sudanian woodland , with 26% of the area in agricultural land use. Upland soils are shallow and typically over laterite. On the average the annual precipitation ranges between 800 and 1200 mm. The population of the department was estimated to be 127,068 in the 1988 census, with a density of 28.6 / km^2, relatively low compared to many parts of Senegal.

The department faces a number of critical land use issues, including the effects of charcoal and fuelwood production, cultivation of cash crops (peanuts and cotton), excessive burning, and agricultural expansion.

The issue most related to absolute change in land use/land cover in Velingara is the tradeoff between the need for keeping the land in woody cover and the need for agricultural land. Agricultural expansion usually occurs at the expense of forested land in this department.

METHODS

Land use/land cover classifications were derived from Landsat MSS images from 1973, 1978, and 1990. The resulting 15 classes were aggregated into 5 major classes (Fig. 2). The resulting vector coverage was gridded at 80 meters to echo MSS resolution. It should be noted, however, that parallel studies were carried out to determine the effects of different resolutions on the Markov and landscape metrics .

Using post-classification differencing, a number of change datasets were developed. The change data was then analyzed using two modified Fortran programs, the temporal program from Harbaugh and Bonham-Carter (1970), with the spatial version coming from Lin and Harbaugh (1984).

Markov Chain Models

The underlying tenet of a 1st order Markov process is based on the probability that the system will be in a given state (land class) at some time t_2 is deduced from the knowledge of its state at time t_1. Therefore, the probability does not depend on the history of the system before time t_1. When a Markov process moves from one time step to the next, the transition from one state to the next only depends on that given state and not on how the process has arrived in that state. In other words, history plays no role in the future. (Parzen, 1962)

For the scope of this work, the Markov process will be considered to have discrete states in the form of five classes of land cover and transitions occurring at discrete times. Also, a transition from one state to another can be thought of in either time or space in which space assumes the role of a discrete event or an individual pixel.

A Markov process is formally described by the transition probability function $P(t|x,t_0)$ which represents the conditional probability that the state of the system will be at time *t*, given that at time t_0 ($< t$) the system is in state *x*. So the transition probability matrix describes the specific character of the system where the elements of the matrix are the individual transition probabilities of one state moving to another state after one time or space increment. The transition matrix is as follows:

$$P = \begin{array}{c} \\ \text{From} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \text{To} \\ \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \end{array}$$

subject to:

$$\sum_{j=1}^{m} p_{ij} = 1 \qquad i = 1, 2, \ldots m$$

A transition probability (p_{ij}) is then the probability that the class *x* will be in state *j* at time *t+1* given it was in state *i* at time *t*.

$$P_{ij}^{t+1} = \Pr [x_{t+1} = j \mid x_t = i]$$

The transition matrix is a powerful analytical tool in itself. It provides a method of probabilistically describing a succession of events in space or time. (Harbaugh and Bonham-Carter, 1970)

Calculation of Transition Probabilities

Transition probabilities are calculated based on the frequency distribution of the observations. Given the assigned land cover classes, a frequency table is developed where a count is made of the transition from one state to another over the specified increment. For example, a count is made of the number of times that forest land cover changes to agriculture for the whole scene from one time period to the next or in space from one grid square to the next. This procedure is continued for all classes and increments that have been collected. When completed the frequency table in each row is summed and the values in each matrix element or transition state are divided by the row sums to compute the transition probability values (Table 1). In each row, the probability values should sum to 1.0. The diagonal of the transition probability represent the self-replacement probabilities where as the off diagonal values indicate the probability of change occurring from one state to another state or class. When each of the row totals are divided by the total number of transitions the marginal probability for each row or class is obtained. The individual marginal probabilities indicate the relative proportion of each state/class at the starting point. Also if the all the rows are identical to the marginal probabilities, the process is independent and therefore non-Markov.

Markov Model Assumptions

There are a number of assumptions that must be met if a Markov modeling approach is meaningful. Listed below are three assumptions that should be addressed, two of which are critical from a mathematical standpoint whereas the third one needs to be tested for but does not compromise the approach.

1) independence/randomness

2) stationarity/homogeneity

3) order of the Markov process (Collins, 1975)

The inherent proposition of a Markov process is that there is some memory from one increment to the next but only from that last state. Where no memory exists, the process is independent and not Markovian. A statistical test of independence can be applied to test the null hypothesis. If the null hypothesis is accepted then the process is independent. If it is rejected then the process may form a first order Markov process. The test statistic is a *log likelihood ratio criterion*. The test statistic is:

$$-2\ln\lambda = 2\sum_{ij}^{m} n_{ij} \ln\left(\frac{p_{ij}}{p_i}\right)$$

p_{ij} = transition probabilities n_{ij} = transition frequencies

p_j = marginal probabilities of the j^{th} column m = number of states

where -2 ln [lambda] is distributed as $[chi]^2$ with where $(m-1)^2$ degrees of freedom. (Harbaugh and Bonham-Carter, 1970) This procedure with minor modifications can be used to test assumptions two and three.

RESULTS

Temporal Markov

The three matrices in Table 1 represent the transition probability values. The null hypothesis for independence was rejected, indicating that the process is Markovian. The probabilities at different steps do not, however, exhibit stationarity and the log likelihood ratio test for stationarity supports this conclusion through the acceptance of the null hypothesis. Intuitively, one can see this from looking at the transitions probabilities in both the 73-78 matrix and the 78-90 matrix, which have not remained constant. For example, the probabilities of agriculture in 1973 remaining in agriculture or transitioning into upland woodland in 1978 are 95% and 5% respectively. The same transitions for 1978 to 1990 are 74% and 24% respectively. Probabilities in the 1973-90 matrix resemble those of the 78-90 matrix.

These results suggest that some influence was present after 1978 that altered the land use/land cover change transition probabilities, though what that is has not been determined. It is possible that it is not entirely a land use phenomenon and that the unequal length of the time steps influences the results. A 1985 image is being acquired and classified to add an additional step and to standardize the step length.

The marginal probabilities help indicate the percentage of the landscape in each class at the beginning of the transition period , i.e. in 1973 nearly 15% of the landscape was in agriculture, 76% in upland woodland, and 6% in riparian/moist gallery forest. In 1990, those percentages had c
hanged to 19%, 72%, and 6% respectively.

Some practical results are the actual land use/land cover changes for the key classes of agriculture, upland woodland, and gallery / riparian forest. It is interesting to note that while 18% of upland woodland was converted to agriculture between 1973 and 1990, with 73% remaining the same, 25% of the land in agriculture was returned to woodland or scrub during the same period with 81% remaining the same.

	1973 - 1990					**1978 - 1990**				
	A	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
A	.7355	.2508	.0129	.0000	.0008	.1499	.8453	.0048	.0000	.0000
B	.1808	.8139	.0053	.0000	.0000	.7444	.2404	.0145	.0000	.0007
C	.0833	.0266	.8901	.0000	.0000	.0770	.0259	.8971	.0000	.0000
D	.0000	.0000	.0000	1.000	.0000	.0000	.0000	.0000	1.000	.0000
E	.0000	.0000	.0000	.0000	.9999	.0000	.0000	.0000	.0000	1.000

	1973 - 1990				
	A	**B**	**C**	**D**	**E**
A	.9543	.0534	.0012	.0000	.0000
B	.0594	.9406	.0000	.0000	.0000
C	.0151	.0000	.9848	.0000	.0000
D	.0000	.0000	.0000	1.000	.0000
E	.0000	.0000	.0000	.0000	1.000

A = agriculture **B** = upland woodland **C** = riparian/moist gallery forest
D = shrub savanna **E** = towns

Table 1. Transition probability matrices .

Spatial Markov

One of the shortcomings of the temporal Markov approach is its inability to account for the spatial context of a grid cell or patch in characterizing transition. Therefore, a spatial Markov approach was developed to determine if it better characterized change in the system. The frequency of class to class transitions between the grid cells were tallied for each pair of adjacent cells using the rook sampling pattern (4 neighbor).

The results show strong spatial autocorrelation as indicated by the extremely high frequencies and probabilities in the diagonals as opposed to the off-diagonals (Table 2). By looking at the off-diagonals in the matrices, one can determine the predominate adjacencies for each class. For example, there is notable adjacency between agriculture and upland woodland, but little between agriculture and the riparian/gallery forest which are in the fossil river valleys, though this conclusion is based on a very limited sample for the latter.

These results, however, are obviously influenced by grain size. The effects of different grain sizes on the spatial Markov transitions were calculated in order to better understand their influence. Resoultions ranging beginning with 40 m. and increasing in powers of 2 through 640 m., with an additional test at 1080 m. were used to determine differences in the results of the transition probabilities. There was no appreciable difference through 640m. At 1080m , the transition to "same state" increased, but only by a few hundredths percent with comparable decrease in the off diagonals.

	A	B	C	D	E
A	.957	.039	.003	.000	.000
B	.014	.977	.007	.002	.000
C	.019	.066	.914	.002	.000
D	.000	.011	.000	.989	.000
E	.074	.000	.000	.000	.926

A = agriculture **B** = upland woodland **C** = riparian/moist gallery forest

D = shrub savanna **E** = towns

Table 2. Spatial Markov transition probabilities.

Year	Class A-B	# of Patches A B		Probability A B		Exp. Joins	Obs. Joins	Variance	Z Score
1973	1-2	132	26	.63	.09	31	137	631.88	4.21
*	1-3	129	34	.63	.17	58	54	906.56	-0.16
1978	1-2	132	26	.61	.12	46	149	766.22	3.70
*	1-3	132	36	.61	.17	64	69	901.67	0.15
1990	1-2	147	47	.57	.18	80	184	930.21	3.40
*	1-3	147	39	.57	.15	67	85	859.95	0.63

1 = agriculture **2** = upland woodland **3** = riparian / moist gallery forest

Table 3. Join-count statistics.

Join-Count Statistics

As a way of comparing the results of the Markov models, the data were processed using other spatial analysis techniques. One such technique, join-count statistics, measures the number and type of joins between adjacent polygons on a map. The frequency of the joins yields information about clustering and spatial autocorrelation (Goodchild, 1986; Bonham Carter, 1994). In order to further investigate the spatial influence of these transitions and to eliminate the problems of grain, join-count statistics were derived for a vector coverage of the land use/land cover (Table 3). Results were generated for the 5 aggregated classes.

Table 3 shows the transitions between classes, with their respective number of patches and probabilities. Expected joins are the number expected if the distribution of patches was random. If the number of observed joins differ significantly from expected, we have what amounts to a proximity index. For example the agriculture - upland woodland transitions are significant as indicated by the corresponding Z scores. A strong proximity or spatial adjacency exists between agriculture and upland woodland. The join-counts indicate there is no significant relationship between agriculture and riparian/gallery forest. These results generally reflect those obtained using the spatial Markov approach.

Landscape metrics

Landscape metrics have been used by landscape ecologists to quantify attributes of spatial patterns (Flam and Turner, 1984; Wickham and Ritters, 1995). The landscape metrics for this study were derived using the program Fragstats (McGarigal and Marks, 1995).

Though just a preliminary examination, a few of the landscape metrics were of particular interest. The most notable outcome was our ability to quantify increased fragmentation as shown by the positive trend in the level of the number of patches and edge density (ratio of edge to area) and the decrease in the mean patch size in both the landscape and individual class metrics (Table 4). There is utility in using these metrics for comparing different landscapes predicted by the modeling process. Additional work is required to determine the relationship between these metrics and the results form the Markov models. The effect of gridding resolution is also unclear, though Wickham and Ritters (1995) suggest there is little effect on the metrics in changing resolutions up to 80 meters.

	Mean Patch Size				**Patch Size Coeff. Var.**				**Edge Density**		
	1973	**1978**	**1990**		**1973**	**1978**	**1990**		**1973**	**1978**	**1990**
A	642.0	783.2	955.8	*	181.8	186.1	271.2	*	3.6	4.17	5.6
B	24,367.3	15,764.8	7,786.3	*	306.8	367.6	504.3	*	5.9	6.4	7.6
C	877.9	827.0	55.0	*	288.3	296.3	249.6	*	2.9	2.9	2.8

A = agriculture **B** = upland woodland **C** = riparian / moist gallery forest

---

Table 4. Landscape metrics

DISCUSSION

This transition of land use/land cover to and from agricultural use is of particular interest (Fig. 3). Based on data collected from a rapid rural appraisal and other interviews conducted in the department by project staff, certain land use trends began to surface. The farmers in Velingara are still maintaining a traditional fallow period up to six or more years in some instances. Part of the shift in and out of agriculture might be explained by this practice. The addition of a 1985 image should help to confirm this assertion. There is some indication, however, that because forest area is still abundant, it is being cleared rather than returning to fallowed land because of the relatively higher fertility (Freudenberger 1996). There is also indication that some clearing is occurring in forest areas to maintain the appearance of "les mettre en valeur" or "productive use", a key requirement of the Land Law to demonstrate one's right to claim user privileges on that land. These are tenure issues that are key to policy makers and managers in the region.

As mentioned, Markov models are mathematically compact, easy to implement with empirical data, and lend themselves well to simulation. But , as to be expected, they have their weaknesses. They operate under fairly restrictive assumptions such as stationarity and the temporal process does not account for spatial context. Results indicate that there has been notable change in the study area and that the frequency of this change has not remained constant, suggesting that landscape change is not strictly Markovian (Turner 1988) . The temporal Markov models capture that aspect of the change well, but changes to the spatial Markov model are necessary to make it as effective.

Both appear to have potential for the simulation of land use/and cover change. The bounds of the temporal and spatial Markov results can be used to define simulations. Ideally we will be able to capture the essence of the change through modeling based on the Markov properties of the landscape. The actual implementation, however, may require modification of the transition probability matrices. Defining this modification is the focus of future research.

FUTURE RESEARCH

In order to address the problems with meeting the assumptions of the first order Markov process, we are investigating Markov variations such as the semi-Markov model (Rogerson 1978, Acevedo 1995) which relaxes the stationarity assumption, or multiple order Markov models which carry memory of the system beyond one step. Another variation is to start with these initial matrices and modify them to increase their utility using a suite of techniques.

We are beginning to address the issue of spatial influence and neighborhood effects . This will include investigation of spatial weighting and is role in developing the spatial transition probability matrix. An attempt is being made to couple the temporal with the spatial models.

In an effort to assess the benefit of integrating socioeconomic and other categorical ancillary data, we are applying logistic regression techniques. Results from this approach will be compared to those from the simple Markov and modified Markov models. They will also be considered as input in the modification of the matrices and as input for a knowledge based approach.

Finally, we feel it is key to integrate both expert knowledge and the influence of socioeconomic factors such as policy reform, etc. not easily captured in the previously described modeling approaches.,e.g. using Markov and regression techniques. Therefore, we hope to develop a knowledge based or expert system that will better accomodate these data.

REFERENCES

Acevedo, M.F., D.L.Urban, and M.Alban, 1995. Transition and Gap Models of Forest Dynamics. Ecological Applications, 5(4):1040-1055.

Baker, W., 1989. A review of models of landscape change. Landscape Ecology, 2(2):111-133.

Bell, E.J, 1974. Markov analysis of Land Use Change - An Application of Stochastic Processes to Remotely Sensed Data. Journal of Socieoeconomic Planning Sciences. 8:311-316.

Boerner, R.E., M.N.DeMaers, J.W.Simpson, F.J.Artigas, A.Silva, and L.A.Berns, 1996. Markov Models of Inertia and Dynamism on Two Contiguous Ohio Landscapes. Geographical Analysis, 28(1):56-66.

Bonham-Carter, G.F., 1994. Geographic Information Systems for Geoscientists : Modelling with GIS. Pergamon. New York.

Bourne, L.S. , 1976. Monitoring Change and Evaluatin the Impact of Planning Policy on Urban Structure : A Markov Chain Experiment. Plan Canada. 16(1):514.

Collins ,L., 1975. An Introduction to Markov Chain Analysis. CATMOG. No.1. Institute of British Geographers. Norwich.

de Graff, J.,1993. Soil Conservation and Sustainable Land Use. Royal Tropical Institute. Amsterdam.

Flam, R.O. and M.G.Turner, 1994. Alternative mode formulation for a stochastic simulation of landscape change. Landscape Ecology, 19(1):37-46.

Freudenberger, K.S., 1996. Kandia: Threatened Forest, Threatened Livelihood. Report, USAID/Dakar. (In press).

Gilruth, P.T., S.E.Marsh, and R.Itami, 1995. A Dynamic spatial Model of Shifting Cultivation in the Highlands of Guinea, West Africa. Ecological Modelling, 79:179-197 (May).

Goodchild, M.F., 1986. Spatial Autocorrelation. CATMOG 47. Institute of British Geographers. Norwich.

Harbaugh, J.W. and G.Bonham-Carter, 1970. Computer Simulation in Geology. Wiley Interscience. New York.

Horn, H.S., 1975. Markovian Properties of Forest Succession. In M.L.Cody and J.M.Diamond,eds. Ecology and Evolution of Communities . Belknap Press, Cambridge, Mass.

Lake, L., 1979. L'expansion du bassin arachidier : Senegal 1954-1979. IFAN, Dakar.

Lin, C. and J.W.Harbaugh, 1984. Graphic Display of Two and Three Dimensional Markov Computer Models in Geology. Von Nostrand Reinhold Co., New York.

McGarigal, K. and B.J.Marks, 1995. Fragstats: Spatial Pattern Analysis Program for Quantifying Landscape Structure. Gen.Tech.Rep. PNW-GTR-351. USDA , Portland, OR.

Parzen, E., 1962. Stocahastic Processes. Holden-Day, Inc., San Francisco.

Rogerson, P., 1978. Extending the Markov Model : The Use of Differential Stationarity. Discussion Paper Series. University of Toronto.

Turner, M.G. , 1990. Landscpae Changes in Nine Rural Counties in Georgia. PERS, 56(3):379-386.

Turner, M.G., 1988. A Spatial Simulation Model of Land Use Changes in a Piedmont County in Georgia. Applied Mathematics and Computation, 27:39-51.

Wasson, R., 1987. Detection and Measurement of Land Degradation Processes. In A. Chisholm and R. Dumsday, Eds., Land Degradation : Problems and Policies. Cambridge University Press, Cambridge.

Wickham, J.D. and K.H.Ritters, 1995. Sensitivity of landscape metrics to pixel size. International Journal of Remote Sensing, 16(18):3585-3594.