**Title**
Workshop on Status and Trends in Spatial Analysis

**Author**
National Center for Geographic Information and Analysis

**Publication Date**
1998-12-01

# WORKSHOP ON STATUS AND TRENDS IN SPATIAL ANALYSIS

Santa Barbara, California: December 10-12, 1998

## Introduction

The National Science Foundation sponsored a workshop from December 10-12, 1998 at the Upham Hotel, Santa Barbara. The workshop included about 35 international scholars, and was organized under the Varenius program, the National Center for Geographic Information and Analysis (NCGIA)'s project to advance geographic information science, by a steering committee of Michael Goodchild (NCGIA/UCSB), Luc Anselin (University of Texas at Dallas), Arthur Getis (San Diego State University), Ayse Can (Fannie Mae), John Paul Jones II (University of Kentucky), Morton O'Kelly (Ohio State University), John Wilson (University of Southern California), and Paul Longley (University of Bristol). Funds were available to support participation.

*Purpose of the meeting:*

Since their inception, geographic information systems have been promoted as vehicles for conducting spatial analysis, that is, for supporting scientists trying to extract meaning and insight from geographic data. Geographers in particular have hoped that GIS would be the 'trojan horse' encouraging and facilitating greater attention to spatial perspectives in other disciplines, and thus raising the utility and practical relevance of geography, and to a large degree this expectation has been realized. But the pace of methodological change in both GIS and spatial analysis has been so rapid in recent years that a stock-taking is appropriate. An assessment is needed of how successful GIS has been at making spatial analysis widely available to physical and social scientists, and of what new directions might be researched in the future. How satisfactory is the environment currently provided by GIS, and how might it be improved? Is spatial analysis being neglected by the sheer diversity of current research in geographic information science? Have GIS and spatial analysis responded appropriately to the critiques published in recent years by social theorists and humanist geographers? How likely are current research efforts to provide an optimum environment for research in geography, regional science, and other disciplines that study the Earth's surface in the coming decade?

This document includes a list of the participants involved in the workshop, as well as the position papers they submitted with their application.

## Participants

Natalia Andrienko
German National Research Center for Information Technology, Sankt-Augustin, Germany

Luc Anselin
Department of Geography, University of Texas - Dallas

Bernie Bauer
Geography, National Science Foundation, Washington, D.C.

Ling Bian
Department of Geography, State University of New York, Buffalo

Nancy Bockstael
Department of Economics, University of Maryland College Park

Barry Boots
Department of Geography, Wilfrid Laurier University

Paul Box
Department of Geography and Earth Resources, Utah State University

Lawrence Brown
Department of Geography, The Ohio State University, Columbus

Aileen Buckley
Department of Geography, University of Oregon, Eugene

Peter Burrough
Department of Geography, University of Utrecht

Ayse Can
Fannie Mae

Diane Cook
Department of Statistics, Iowa State University

Noel Cressie
Department of Statistics, Iowa State University (until June 1999)
Department of Statistics, The Ohio State University (after June 1999)

Jack Dangermond
Environmental Systems Research Institute, Redlands, California

Marcus Felson
Department of Sociology, Rutgers University

A. Stewart Fotheringham
Department of Geography, University of Newcastle, United Kingdom

Andrew Frank
Department of Geoinformation, Technical University of Vienna,  Austria

Janet Franklin
Department of Geography, San Diego State University

Mark Gahegan
Department of Geographic Information Science, Curtin University of Technology, Perth, Western Australia

Jacqueline Geoghegan
Department of Economics, Clark University,Worcester, MA

Art Getis
Department of Geography, San Diego State University

Michael Goodchild
NCGIA, University of California-Santa Barbara

Daniel A. Griffith
Department of Geography, Syracuse University, New York

Steve Guptill
U. S. Geological Survey, Reston, Virginia

Bob Haining
Department of Geography, University of Sheffield, Sheffield, UK

Geoffrey M. Jacquez
BioMedWare, Ann Arbor, Michigan

John Paul Jones
Department of Geography, University of Kentucky

Konstantin Krivoruchko
Environmental Systems Research Institute, Redlands, California

Mei Po Kwan
Department of Geography, The Ohio State University

Tony Lea
Micromarketing Data and Systems, Toronto, Canada

Michael Leitner
Department of Geography and Anthropology, Louisiana State University, Baton Rouge

Yee Leung
Department of Geography, Chinese University, Hong Kong

Paul Longley
Department of Geography, University of Bristol

Duane Marble
Center for Mapping & the Department of Geography, The Ohio State University, Columbus

Harvey Miller
Department of Geography, University of Utah, Salt Lake City

Atsuyuki Okabe
Department of Urban Engineering, University of Tokyo, Tokyo

Morton O'Kelly
Department of Geography, The Ohio State University, Columbus

Pat Pellegrini
Department of Geography, The Ohio State University, Columbus

Jonathan D. Phillips
Department of Geography, College of Geosciences, Texas A&M University, College Station

Rhonda Ryznar
Department of City and Regional Planning, University of North Carolina

Dan Sui
Department of Geography, Texas A&M University, College Station

Lance Waller
Department of Biostatisitcs, University of Minnesota

Elizabeth A. Wentz
Department of Geography, Arizona State University, Tempe

John Wilson
Department of Geography, University of Southern California

## Position Papers

In Alphabetical Order

**Natalia V. Andrienko**
German National Research Center for Information Technology, Sankt-Augustin, Germany

### Knowledge Extraction from Spatially Referenced Databases: a Project of an Integrated Environment

#### Current state

The notion of **K**nowledge **D**iscovery in **D**atabases (KDD) denotes the work on revealing significant relationships and regularities in data based on the use of algorithms collectively entitled “data mining”. The KDD process consists in an iterative fulfillment of the following steps [5]:

1. Data selection and preprocessing, such as checking for errors, removing outliers, handling missing values, and transformation of formats.
2. Data transformations, for example, discretization of variables or production of derived variables.
3. Selection of a data mining method and adjustment of its parameters.
4. Data mining, i.e. application of the selected method.
5. Interpretation and evaluation of the results.

In this process the phase of data mining takes no more than 20% of the total workload. However, this phase is much better supported methodologically and by software than all others [6]. This is not surprising because performing of these other steps is a matter of art rather than a routine allowing automation [7]. Lately some efforts in the KDD field have been directed towards intelligent support to the data mining process, in particular, assistance in the selection of an analysis method depending on data characteristics [2,3].

A particular case of KDD is knowledge extraction from spatially referenced data, i.e. data referring to geographic objects or locations or parts of a territory division. In analysis of such data it is very important to account for the spatial component (relative positions, adjacency, distances, directions etc.). However, information about spatial relationships is very difficult to represent in discrete, symbolic form required for the data mining methods. Known are works on spatial clustering [4] and use of spatial predicates [8], but a complexity of data description and large computational expenses are characteristic for them.

**Our suggestion**

For the case of analysis of spatially referenced data we propose to integrate traditional data mining instruments with automated cartographic visualization and tools for interactive manipulation of graphical displays. The essence of the idea is that an analyst can view both source data and results of data mining in the form of maps that convey spatial information to a human in a natural way. This offers at least a partial solution to the challenges caused by spatially referenced data: the analyst can easily see spatial relationships and patterns that are inaccessible for a computer, at least on the present stage of development. In addition, on the ground of such an integration various KDD steps can be significantly supported.

The most evident use of cartographic visualization is in evaluation and interpretation of data mining results. However, maps can be helpful also in other activities. For example, visual analysis of spatial distributions of different data components can help in selection of representative variables for data mining and, possibly, suggest which derived variables would be useful to produce. On the stage of data preprocessing a map presentation can expose ‘strange’ values that may be errors in the data or outliers. Discretization, i.e. transformation of a continuous numeric variable into one with a limited number of values by means of classification, can be aptly supported by a dynamic map display showing spatial distribution of the classes. With such a support the analyst can adjust the number of classes and class boundaries so that interpretable spatial patterns arise.

More specifically, we propose to build an integrated KDD environment on the basis of two existing systems, **Kepler** [9] for data mining and **Descartes** [1] for interactive visual analysis of spatially referenced data. **Kepler** includes a number of data mining methods and, what is very important, provides a universal plug-in interface for adding new methods. Besides, the system contains some tools for data and formats transformation and is capable of graphical presentations of some kinds of data mining results (trees, rules, and groups). **Descartes** automates generation of maps presenting user-selected data and supports various interactive manipulations of map displays that can help to visually reveal important features of spatial distribution of the data. **Descartes** also supports some data transformations productive for visual analysis. It is essential that both systems are designed to serve the same goal: help to get knowledge about data. They propose different instruments that can complement each other and together produce a synergistic effect.

In its present state, **Kepler** contains the following data mining methods:

1. Methods **fw** and **kNN** estimate importance of different variables in relation to values of a selected variable.
2. Methods **C4.5** and **C5.0** derive classification trees.
3. Methods **C4.5**, **FOIL**, and **BNGE** generate classification or prediction rules.
4. Methods **SIDOS** and **MIDOS** find statistically interesting subgroups of objects with regard to distribution of values of a variable.
5. Method **AutoClass** performs clustering.

Most of the methods (groups 1-4) require selection of a target variable that typically should be discrete and are intended for revealing relationships between the target variable and other variables selected for the analysis. **Descartes** can be effectively used for producing 'promising' discrete variables including, implicitly or explicitly, a spatial component. The following ways of doing this are available:

1.Classification by segmentation of a value range of a numeric variable into subintervals.
2.Cross-classification of a pair of numeric attribute. In both cases the process of classification is highly interactive and supported by a map presentation of the spatial distribution of the classes that reflects in real time all changes in the definition of classes.
3.Spatial aggregation of objects performed by the user through the map interface. Results of such an aggregation can be represented by a discrete variable. For example, the user can divide city districts into 'center' and 'periphery' or encircle several regions, and the system will generate a variable indicating to which aggregate each object belongs.

Results of most of the data mining methods are naturally presentable on maps. The most evident is the presentation of subgroups or clusters: belonging of a geographical object to a subgroup or a cluster can be designated by painting or an icon. The same technique can be applied for tree nodes and rules: visual features of an object indicate whether it is included in the class corresponding to a selected tree node, or whether a given rule applies to the object and, if so, whether it is correctly classified.

Since **Kepler** contains its own facilities for presentation (non-geographical) of data mining results, it would be productive to make a dynamic link between **Kepler**'s and **Descartes**' displays. This means that, when a cursor is positioned on an icon symbolizing a subgroup, a tree node, or a rule in a **Kepler**'s display, the corresponding objects are highlighted in a **Descartes**' map. And vice versa, selection of a geographical object or a group of objects in a map results in highlighting subgroup(s) or tree nodes it belongs (or they belong) to or rules applicable to it (them).

Besides their main capabilities (data mining in **Kepler** and data visualization plus analysis-supporting display manipulation in **Descartes**), the systems contain additional useful functions and components to be included in the integrated environment. Thus, **Kepler** contains a tool **DataZoom** [10] supporting analysis of tables with data by a highly interactive dynamic interface for sorting, focusing, and querying. **Kepler** can also perform a number of necessary routine operations over datasets: transformations of formats, access to databases, querying etc. Descartes has a convenient graphical interface for outlier removal and an easy-to-use tool for generation of derived variables by means of arithmetic operations over existing variables.

The above-presented consideration can be summarized in the form of three kinds of links between data mining and cartographic visualization:

From geography to mathematics: using dynamic maps, the user arrives at some geographically interpretable results or hypotheses and then tries to find an explanation of the results or checks the hypotheses by means of data mining methods.

From mathematics to geography: data mining methods produce results that are then visually analyzed after being presented on maps.

Linked displays: graphics representing results of data mining in the usual (non-cartographic) form are viewed in parallel with maps, and dynamic highlighting visually connects corresponding elements in both types of displays.

### Software implementation

The feasibility of software implementation of the project is supported by the circumstance that both systems have a client-server architecture and use the TCP/IP protocol for the client-server communication. The client components of both systems are realized in the Java language.

For coupling the two systems, it is necessary to organize their shared use of the same data and to create a mechanism to distribute and transfer control between the systems. For this purpose a communication protocol should be designed and implemented.

### References

Andrienko, G. and Andrienko, N. (1998) Intelligent Visualization and Dynamic Manipulation: Two Complementary Instruments to Support Data Exploration with GIS. In Proceedings of AVI'98: Advanced Visual Interfaces Int. Working Conference (L'Aquila Italy, May 24-27, 1998), ACM Press, pp.66-75.

Brodley, C. (1993) Addressing the Selective Superiority Problem: Automatic Algorithm / Model Class Selection. In Machine Learning: Proceedings of the 10th International

Conference, University of Massachusetts, Amherst, June 27-29, 1993. San Mateo, Calif. : Morgan Kaufmann, pp.17-24.

Gama, J. and Brazdil., P. (1995) Characterization of Classification Algorithms. In Progress in Artificial Intelligence, LNAI 990, Berlin: Springer-Verlag, pp.189-200.

Gebhardt, F. (1997) Finding Spatial Clusters. In Principles of Data Mining and Knowledge Discovery PKDD 97, LNCS 1263, Berlin: Springer-Verlag, pp.277-287.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996) The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM, 39 (11), 27-34.

John, G.H. (1997) Enhancements to the Data Mining Process. PhD dissertation, Stanford University.

Kodratoff, Y.. (1997) From the art of KDD to the science of KDD. Research report 1096, Universite de Paris-sud.

Koperski, K., Han, J., and Stefanovic, N. (1998) An Efficient Two-Step Method for Classification of Spatial Data. In Proceedings SDH 98, Vancouver, Canada: International Geographical Union, pp.45-54.

Wrobel, S., Wettschereck, D., Sommer, E., and Emde, W. (1996) Extensibility in Data Mining Systems. In Proceedings of KDD 96 2nd International Conference on Knowledge Discovery and Data Mining. AAAI Press, pp.214-219.

Spenke, M., Beilken, Chr., and Berlage, Th. (1996). The Interactive Table for Product Comparison and Selection. In Proceedings of the UIST 96 9th Annual Symposium on User Interface Software and Technology, Seattle, November 6 - 8, 1996. ACM Press, pp.41-50.

**Luc Anselin**
Bruton School of Development Studies
University of Texas at Dallas

## GIS, Spatial Econometrics and Social Science Research

The subset of the domain of spatial analysis that pertains to the statistical analysis of spatially referenced data has recently gained a growing acceptance as a methodology in the mainstream social sciences. I will focus my remarks on this specific issue, leaving the discussion of aspects of spatial analysis such as optimization and decision support systems to others.

The recent dissemination of a spatial analytic perspective in the social sciences (*outside* of the discipline of geography) is often attributed to the rapid spread of GIS technology to the desktop and the availability of a vast array of geographically referenced socio-economic data. This has led to the use of GIS for data organization and visualization as well as increasingly in an inductive approach to exploring data for meaningful patterns and structures (exploratory spatial data analysis). While these have undeniably been important factors, an equally crucial aspect has been the need to operationalize 'new' theoretical constructs that explicitly incorporate space in the analysis of human (economic) behavior. Many of these concepts are similar (though not always acknowledged) to the models proposed by economic geographers and regional scientists in the 1960s, and stress the importance of location, neighborhood, region and spatial (social) interaction. Current examples in economics are the emphasis on spatial externalities and regional clusters (e.g., Krugman, Arthur, Porter), theories of interacting agents and interdependent decision making (e.g., Pollak, Ioannides, Durlauf, Brock, Brueckner), the importance of social interaction and group effects (e.g., Akerlof, Aoki) and neighborhood effects (Borjas). Similar examples can be cited in recent work in other social sciences, such as sociology, political science and criminology. Unlike their antecedents of the 1960s, description and discussion of these theories appears in the core journals of the mainstream disciplines, such as the *American Economic Review, Journal of Political Economy and Econometrica* for economics.

Empirical validation of the new spatial concepts and models requires an explicit spatial econometric methodology that tackles issues of spatial dependence and spatial heterogeneity, as well as their extensions in the space-time domain. Spatial econometrics is a subset of spatial statistics in that rather than being statistics for (any) spatial data, it concerns itself with statistics for spatial (socio-economic) *models*, where the model specification is dictated by theory. These subtle differences aside, it is important to acknowledge that a growing number of mainstream econometricians (e.g., Kelejian, Prucha, Bera, Baltagi, Pinkse) have started to contribute to the spatial econometric methodology and that spatial econometrics has gained recognition as a useful subset of the econometric toolbox.

These recent developments in the mainstream social sciences in general and in economics in particular raise a number of challenges for the next generation of spatial analysis. Central to this is the need to move beyond mapping (generally recognized in the GIS community, but not necessarily in the mainstream disciplines) and to tackle the methodological and theoretical issues that address the complexities of the current models. I see the potential for new developments in three important domains:

- *Extending concepts of space*

Spatial analysis needs to go beyond dealing with physical geographical locations to include location in 'social' space (social distance, economic distance). This will require further consideration and development of distance metrics for 'social' space, for space-time dynamics and notions of 'topology' in space-time (the counterpart of the 'weights' matrix in spatial autocorrelation analysis). Promising avenues are current work on GIS data models, object-oriented GIS, and the like.

- *Broadening the analytical toolbox*

The toolbox of spatial econometrics and spatial analysis needs to be extended to deal with the challenges posed by the analysis of socio-economic space-time data. While much progress has been made, some unresolved issues are the estimation of space-time dynamics for limited dependent data (such as discrete choice data, duration data), modeling changing choice sets, distinguishing spatial dependence from spatial heterogeneity, effective visualization of model fit, etc. For many of these research questions analytical solutions are impossible or prohibitive, such that computational approaches must be followed (e.g., simulated moments, simulated likelihood, Markov Chain Monte Carlo). This requires advances in computational geography in the form of the development of new and/or efficient algorithms to tackle the complexity of realistically sized data sets

- *Technology transfer*

Most of the current commercial GIS software comes in the form of (partially) open environments that allow the user to include customizations and extend the functionality. In a modern component oriented computing environment, there is therefore no longer a high priority to have commercial spatial analytical tools included in the 'box', but rather to have the mechanisms to mix and match components to accomplish specific tasks. Since the commercial world will always be behind the curve when it comes to 'state of the art' in terms of the statistical methodology it delivers, such a toolbox (such as MapObjects) allows analysts to integrate their own selection of analytical methods with core GIS functionality. In contrast to the toolbox approach, shrink-wrapped commercial GIS software has tended to offer the lowest common denominator when it comes to spatial analytical (let alone statistical) methodology. In my view, this has had two major drawbacks. One is that uninitiated users identify 'spatial analysis' with the (limited) set of techniques offered by a software vendor. The other is that the analysis is presented as being 'easy' and underlying assumptions, algorithms and limitations are hidden from the user. Both issues pose challenges to software developers as well as to current GIS education, both in the academic and in the private sector (by vendors).

The theoretical questions posed in the mainstream social sciences offer an important challenge to the methodology of spatial analysis. However, this also constitutes a major opportunity for the spatial analytical perspective (as part of a geographic information *science*) to contribute to the theoretical debate in the core disciplines.

**Ling Bian**
Department of Geography, State University of New York, Buffalo

## Status of GIS Use in Physical Sciences

It may be necessary first to identify different types of community within GIS and within physical sciences. In GIS, there are the research community, software developers, and GIS professionals. In physical sciences, there are the research community and, in a loose sense, the professionals. These groups are related to each other and play different roles in GIS development. The GIS developers, among all groups, have the most direct impact on physical scientists' work and their perception about GIS. GIS systems (as opposed to GIScience) will be the focus of the following discussion. The "physical scientists" refer to those whose research is on physical processes of natural world, for example, those who work in general areas of hydrology, soils science, atmospheric science, oceanography, forestry, or ecology.

For physical science research, the role of current GIS systems has remained as "front ends" (pre-processing spatial data to prepare model input) and "back ends" (visualizing model output spatially). It has become clear that the analysis functions of GIS cannot replace the analysis functions in process models. The GIS functions are designed for manipulating and extracting information needed for the models. These analytical capabilities of GIS have remained weak in comparison with other capabilities that GISs have promised to deliver, such as spatial data management and visualization. The improvement in analysis capability has not caught up with the pace of other GIS capabilities. This weakness affects physical scientists' ability to deal with spatial phenomena. The often heard "We know everything about GIS the geographers do" from ecologists reflect such limitation in current GISs.

The weak capability of spatial analysis is not the only impedance for physical scientists to use GIS. The better developed data management and visualization capabilities of GIS have not always delivered a satisfactory user environment. For example, data format incompatibility between GIS and models is a simple technical problem, but it is one of the most costly problems for users in modeling community (Karimi, 1997). It is common for modelers to spend much greater proportion of time, energy, and resources on data conversion than on model calibration. Automating the conversion is model-dependent and it does not alleviate the burden for physical scientists who are constantly involved with different models. Although it is unrealistic to expect GISs to provide data format for all models, current GISs have not provided tools to ease the conversion.

Relational database remains to be the dominant framework for storing spatial data despite many academic discussions over its strengths and shortcomings (Kim and Lochovsky, 1989). For models that describe dynamic processes, there are at least three types of relations: (1) relations between different variables at a fixed location, that are represented by the mathematical functions, (2) relations between different locations for a fixed variable, such as in the situation of finite difference for hydrology, and (3) a combination of (1) and (2), in that both locations and variables are related. Relational GIS databases cannot adequately accommodate any of the relations. This makes it difficult for modeling to directly use data manipulation functions built in GIS databases. Conceptual compatibility between GIS and modeling, such as raster data structure with finite difference as well as TIN with finite element, have remained as academic discussions. Current GIS database does not allow direct access to its data structures so that modeling functions can be linked. Developing in-house code is still a much more practical approach to use the raster- or TIN-like data. With the difficulties of data format and database, GIS has retreated to more of a data provider (e.g., DEMs, DLGs) than a data analyzer to many physical scientists.

Efforts have been made to ease the problems. Integrating GIS with environmental models has been the title of three international conferences (Goodchild et al., 1993; Goodchild et al., 1996; NCGIA, 1996). Some early efforts attempted to build GIS functions within a process model or more often models are rebuilt within a GIS (Betty and Xie, 1994). This approach has proved to be limited. More successful or more practical approach has been leaving the GIS and models essentially intact but bridging the two together. Integration strategies such as simple data file transfer, loose coupling, and tight coupling (Chou and Ding, 1992; Nyerges, 1993; Abel et al, 1994) are daily practice in many physical scientific work. Each strategy has its benefits and costs. The more recent development of OpenGIS specifications for spatial data and function (Buehler and McKee, 1996) holds new promises to alleviate the daily burden of integrating. At least it provides a solution for data format incompatibility problem. The aforementioned efforts seek for technical solutions. Semantic compatibility was assumed to be handled by the end user; thus it is rarely discussed under the topic of integration problem.

Scale incompatibility is a problem beyond technical solution. Differences in development history may have contributed to the mismatch. Many process models used today were originally developed in 1970s when computer became available. The popular use of GIS came at least a decade later in mid 1980s. In coping with the lack of means to handle spatial data, the original model development was restricted to simplistic treatments of spatial variation, for example, using coarse spatial resolution or small spatial extent. Many models use raster-like data because it is a simple way to partition the space. This may explain why raster-based GIS packages such as GRASS is so popular among modelers (Being a public domain package and having open architecture are not the only reasons). Not only the model requires coarse resolution input, it also simulates the physical processes that occur at corresponding spatial scales. With these traits carried to today, many models are prohibited from taking the advantage of the details provided by today's GIS data. Often the GIS data must be aggregated to a coarser resolution before they can be entered into a model (Zack and Minnich, 1991). This problem is more inherent than problems such as data format incompatibility.

The representational difference between GIS and process modeling is more challenging. Current GISs represent static, layered world through spatial data models. Process models use mathematical functions to model the dynamics of the world (Maidment, 1993, 1996). For distributed (or spatially explicit) models, raster data structure is still the best available. Current GIS data in general cannot accommodate the need for representing dynamic processes. The object orientation paradigm offers many advantages for this purpose (Raper and Livingstone, 1995), but it is better suited to object-like phenomena. The process models deal primarily with field-like, continuous phenomena. It is conceptually as well as technically difficult for object orientation to implement dynamics of fields. Kemp (1997a, 1997b) addressed the issue of integrating field data and process models from representational perspective. In-depth analysis as such is much needed.

Another difference between GIS and process models lies in the fact that GIS is meant to provide an objective representation of the world through stored measurements and observations. Information may be eventually extracted based on a particular need (Peuquet, 1994). In this sense, GIS is a generalist. In contrast, modeling usually focuses on a particular process; thus process models are specialists. A generalist GIS package cannot always meet requirements of specialist models. This mismatch has also caused the gaps in practical use of GIS in physical sciences.

In addition, effective tools to represent and handle three dimensional data are still not readily available (Scott, 1997). Furthermore, very little has been studied about representing flow in GIS. The two issues are important because most natural processes are three dimensional, and flow of energy and material is the core concept of physical sciences.

The incompatibility is certainly two-sided. Take the data format problem as an example, not only environmental models are incompatible with GISs, but also among the models themselves. Many environmental models are monolithic, legacy models. Developing platform- and language- independent modules (or algorithm library, or component-ware) is a technically feasible solution (Leavesley et al., 1996). Such a solution, however, requires resources. Institutional support is much more critical than technical solutions. Investing effort into such development may not be seen as the right path for career advance for many physical scientists. It is less likely that the commercial software developers will take the task. Research communities in physical science may play an important role in GIS development, but it represents a small and diverse market for GIS products.

Incompatibility between GIS and process models in terms of data format, database, scale, and representation requires different solutions, from simpler technical solutions to more sophisticated representational ones. Some technical solutions are already on the way. Representational solutions will take longer time and more efforts. Above all, institutional solution is always more critical and more difficult to achieve.

## References

Abel, D.J., Kilby, P.J., and Davis, J.R. (1994) The systems integration problem. International Journal of Geographical Information Systems, 8(1): 1-12.

Batty, M. and Xie, Y. (1994) Modeling inside GIS: Part 1: model structures, exploratory spatial data analysis and aggregation. International Journal of Geographical Information Systems, 8(3): 291-307.

Buehler K. and McKee L. (1996) The OpenGIS Guide. http://www.opengis.org, The Open GIS Consortium, Inc.

Chou, H.C., and Ding, Y. (1992) Methodology of integrating spatial analysis/modeling and GIS. Proceedings, 5th International Symposium on Spatial Data handling, Charleston, South Carolina, August 3-7. 514-523.

Goodchild, M. F., B. O. Parks, and L. T. Steyaert (Eds.) Oxford University Press, New York.

Goodchild, M.F., Steyaert, L.T., Parks, B.O., Johnston, C., Maidment, D., and Glendinning S. (Eds.), GIS World, Inc. Fort Collins, Colorado.

Karimi, H.A. (1997) Interoperable GIS applications: tightly coupling environmental models with GISs, Proceedings, International Conference on Interoperating Geographic Information Systems. Santa Barbara, California, December 3-4.

Kemp, K.K. (1997a) Fields as a framework for integrating GIS and environmental process models. Part 1: representing spatial continuity. Transactions in GIS, 1(3): 219-234.

Kemp, K.K. (1997b) Fields as a framework for integrating GIS and environmental process models. Part 2: specifying field variables. Transactions in GIS, 1(3): 219-234.

Kim, W. and Lochovsky, H. (1989) Object-Oriented Concepts, Databases, and Applications. ACM Press, New York.

Leavesley, G.H., Restrepo, P.I., Stannard, L.G., Frankoski, L.A., and Sautins, A.M. (1996) MMS: a modeling framework for multidiscipliary research and operational applications. In GIS and Environmental Modeling: progress and Research Issues. Goodchild, M.F., Steyaert, L.T., Parks, B.O., Johnston, C., Maidment, D., and Glendinning S. (Eds.), GIS World, Inc. Fort Collins, Colorado. 155-163.

Maidment, D. R. (1993) GIS and hydrologic modeling. In Environmental Modeling with GIS, Goodchild, M. F., B. O. Parks, and L. T. Steyaert (Eds.) Oxford University Press, New York. 147-167.

Maidment, D.R. (1996) Environmental modeling with GIS. In GIS and Environmental Modeling: progress and Research Issues. Goodchild, M.F., Steyaert, L.T., Parks, B.O., Johnston, C., Maidment, D., and Glendinning S. (Eds.), GIS World, Inc. Fort Collins, Colorado. 315-323.

NCGIA (1996) Third International Conference/Workshop Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21-25. NCGIA, http://www.ncgia.ucsb.edu.

Scott, M. (1997) Extending map algebra concepts for volumetric geographic analysis. Proceedings, GIS/LIS '97. Cincinnati, Ohio, October 28-30.

Nyerges, T. (1993) Understanding the scope of GIS: its relationship to environmental modeling. In Environmental Modeling with GIS, Goodchild, M. F., B. O. Parks, and L. T. Steyaert (Eds.) Oxford University Press, New York. 75-93.

Peuquet, D.J. (1994) It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems. Annals of the Association of American Geographers, 84(3): 441-461.

Raper, J., and Livingstone, D. (1995) Development of a geomorphological spatial model using object-oriented design. International Journal of Geographical Information Systems, 9(4): 359-383.

Zack, J.A. and Minnich R.A. (1991). Integration of geographic information systems with a diagnostic wind field model for fire management. Forest Science, 37(2): 560-573.

**Nancy E. Bockstael**
Department of Agricultural and Resource Economics, University of Maryland, College Park

Despite the surge in GIS technology and implementation, spatial analysis remains relatively rare in the economics literature. Within economics, the field of environmental and natural resource economics probably stands to gain the most from spatial data and analysis, and it is with respect to this field that the following comments are made.

There are two simple and overarching reasons why economists have pursued little spatial analysis. The first is that spatially explicit economics data is difficult to obtain. The second is that many economists remain unconvinced that space matters in the systems that they study.

While the second problem is fundamental, the first problem is not inconsequential. Satellite imagery, remote sensing technology, and over-flight photography provide a means of registering physical phenomenon, but only rarely are the things that economists study observable in this sense. Economists model human decisions, and only occasionally are these decisions ‘deducible’ from pictures. Land use change is one of the few examples where they are. But even in this case the boundaries of the decision units are unobservable from above, as are many of the explanatory variables that would be needed at a spatially disaggregated level to incorporate into the models of human behavior (e.g. prices, incomes, transactions, etc.)

It is notable that the economic analyses that have succeeded in incorporating spatial data are those where either the human decision is observationally deducible (e.g. deforestation and land use) or where the human activity involves recording location for some other reason (e.g. housing transactions, regulated marine fishing activities). Attempts at collecting data on human actions at a spatially disaggregated level are frustrated, however, by confidentiality regulations that prohibit the dissemination of data that would allow one to deduce the identity of an individual firm. Because geocoding provides location, it inherently violates those regulations. This has presented particular obstacles to spatial analysis of agricultural systems, because agriculture has even more stringent privacy provisions than most industries.

Nonetheless, spatial data of interest to economists is on the increase, and more could be done with spatially explicit data if economists chose to use it in their empirical work. But many economists, even within the environmental and resource economics profession, remain unconvinced of the value of spatial analysis. At best they see geographic information systems as a means of capturing and storing a richer data base or providing greater accuracy and variability in variables for their otherwise aspatial analysis. The analyses that actually care about space, generally do so because space matters to someone else. For example, the location of human activity matters because of its effects on the environment (such as a subwatershed or wetland) or the stock of natural resources (like fish or forests) none of which are themselves locationally fungible.

An altruistic motivation is likely to have a limited impact on a profession like economics.  Only if it matters inherently to the economics of the problem will spatial analysis reach a suitable level of sophistication in economics research.  From a methodological viewpoint, there have been a number of advances, with spatial econometrics methodology for handling spatial data paralleling the advances made decades ago for time series. Yet broad acceptance of the importance of temporally dynamic models of economic processes does not have a parallel in spatially dynamic models.  In fact there are few inherently spatial models in economics.  The ones that come to mind are the monocentric city model of land price gradients and the new economic geography as typified by Krugman's work, some of this arising in the regional economics field.   And there are even fewer empirical applications.  This is despite the fact that there are some inherently spatial processes such as endogenous interactions, contagion, diffusion, dissemination that have interest for economists. Economics will benefit more from efforts (many of which may be underway already) that are devoted to the development of explicitly spatial models of economic processes as well as adaptation of existing spatial models to economics problems from other disciplines.  At this point, the development and dissemination of ways of modeling spatial processes will contribute more to spatial analysis in economics than further attempts at improving accessibility to GIS data and software.

**Barry Boots**
Department of Geography, Wilfrid Laurier University

### How successful has GIS been at making spatial analysis widely available to physical and social scientists?

I would like to focus my comments on this question since I believe the widespread dissemination of spatial analytical procedures through GIS is crucial for the future well-being of not only of spatial analysis (SA) and GIS but also the discipline of Geography. It is thus with considerable disappointment and concern I note that, from my perspective, only very limited success appears to have been achieved. Why is this so?

One reason is something over which spatial analysts and GIS people have little or no direct control. Take a look at recent issues of the "Annals" or "Transactions". Assuming that the contents of these journals accurately reflect contemporary scholarship in geography, it is obvious that many of our colleagues are simply not concerned with explicitly spatial issues whose examination would be enriched by sophisticated spatial analysis. The most that GIS can contribute to such work is to provide a convenient means of data storage and mapping capabilities. In view of this, it becomes imperative that we communicate effectively with those, both inside geography and outside, who still have an interest in the spatial analytical paradigm. How can we do this?

First of all we need to demonstrate more convincingly that GIS based SA techniques can be used to address important substantive problems and that such techniques are capable of providing meaningful insight which cannot be obtained otherwise. Why is it, for example, that despite the development over the past twenty years of diagnostics for spatial dependency, we still routinely encounter published papers which feature mis-specified regression models? Are the authors, reviewers, and editors of such papers unaware of our material or have we failed to convince them of its value? Fortunately, there are also positive signs. For example, interest in the Getis statistic has been increasing in image analysis in remote sensing as it has been shown to capture in a single value much of the same information that required the calculation of a battery of statistics based on co-occurrence matrices. Few individuals combine GIS and SA expertise and even fewer combine this with a topical expertise and so it is not surprising that the illustrations which accompany our presentations of new techniques often appear inconsequential or to reflect more the availability of a convenient data set. I think we need to do much more collaborative work with topical specialists. Not only will this make our techniques more visible in applied arenas, more importantly, it will also help us to identify meaningful spatial questions for which new techniques need to be developed.

However, it is one thing to perk the interest of a few topical specialists, it is another matter to enable our techniques to become part of the everyday tool kits of the majority of their fellow practitioners. This requires that not only should such individuals be aware of our analytical procedures, they must also be able to implement them with minimum effort. The former requires good, practical texts which illustrate both the conceptual underpinnings of our procedures and clear guidance on how to use them. The latter requires user-friendly, integrated software. The optimum would be a combination of the two. As far as I am aware, there is only one example of this (but a good one), Bailey and Gatrell's INFO-MAP. I think that there is a great need to develop software and accompanying texts in the genre the SPSS* manuals. There is no doubt in my mind that the ease of use of SPSS* has contributed to non-spatial statistical procedures becoming commonplace throughout the social sciences. In this regard, I feel we should be a little less elitist and a little more tolerant in the way we present our material. For example, there are some very good spatial statistical texts around but they are not easily accessible to those without formal statistical training (and without either the time or the desire to obtain it). I'm not advocating that we lower our standards but that we change our emphasis, at least as far as textbooks are concerned. We have to recognize that the cultures of other disciplines may differ considerably from our own. The majority of practitioners in other disciplines who can benefit from our materials will only use them if they are packaged in accessible and readily usable forms. There is a large number of potential users for our materials, we cannot afford to ignore them.

## Paul Box

Department of Geography and Earth Resources, Utah State University, Logan

I believe that the most interesting aspects of spatial analysis in the future will be with the ability to model processes at the most local level (individual or cell), and analyze the emergent properties of the coarser scale phenomena. I firmly believe that recent advances both in theory and techniques of artificial life, artificial intelligence, and complex systems theory will provide the framework where many of the next major advances in geography and spatial analysis will occur. Analytical methods involving multi-agent simulations and cellular automata are helping make these modeling frameworks available to the "masses". Toolkits such as the Swarm package, combined with cheap computing power and software support, will provide the critical mass of researchers who will find highly creative ways to approach these problems. The ability for researchers to communicate their results quickly via the Internet will increase the rate at which new discoveries are made, which may force researchers to reevaluate how they publish their results.

One particularly fertile field of research using these techniques will be investigations into situations where behavior of individuals (humans, vehicles, animals, etc.) have impacts on their environment, and the impacted (or changed) environment in turn affects the behavior of the individuals. In such situations, knowledge of processes at one level cannot predict what will happen at a higher level: for example, knowledge of a particular cow's physiology and eating habits will not be enough to explain which part of a field or rangeland will suffer from overgrazing. A similar analogy can be made for human use of the environment.

One problem of the recent advances in artificial life is that they are bringing in issues that are not really "new". Many questions being studied via these methods have already been modeled using established analytical techniques, though often at the expense of impossibly unrealistic assumptions. While the actual contributions of these modeling frameworks to spatial analysis is not clear, there seems to be great promise and considerable excitement.

I am especially interested in attending a forum where people with knowledge and experience in this field can meet and disagree with me.

**Lawrence A. Brown**
Department of Geography, The Ohio State University, Columbus

## THE GIS/SA INTERFACE FOR SUBSTANTIVE RESEARCH(ERS): A CRITICAL NEED

A central question of this conference is "how successful GIS has been at making spatial analysis widely available to physical and social scientists". To put the question somewhat differently, to what degree is there an intersection of GIS and Spatial Analysis which maximizes the power of this combination; facilitates its use by scientists who are focussed on substantive, empirical-based research rather than methodology per se; and propagates such use by example, instruction, and the like?

In my estimation, and speaking as the type of person described in the question, an end-user carrying out empirical research, the answer is "not enough" and the ongoing scenario is not encouraging. Put another way, I believe there is a tremendous need for classroom training, workshops, published examples that can serve as guideposts, and software development -- all of which would facilitate the use of analytic methods that exploit the power of the GIS/SA interface among a broad range of end-users with varying levels of, or inclination towards, methodological expertise.

Addressing this need can make a significant (monumental) difference in the standing of Geography among other sciences. It also can be a significant catalyst towards shifting current research trends/tendencies that too often neglect empirical analysis, particularly of a statistical nature.

Discussion around this broad point can be divided into 6 topics. (1) Parallels with the Quantitative Revolution; (2) Types of GIS Efforts; (3) Gains from the GIS Revolution; (4) GIS/Spatial Analysis in Geography and Available Analytical Tools; (5) Who uses GIS/SA in Scholarly Research that is Substantively Motivated; (6) What is the problem, how can it be remedied? In addressing this topic, I do so as an end-user Geographer concerned with advancing the understanding of academic issues central to our respective disciplines; solving problems by applying GIS/SA tools is an important task, but not among those I identify as central in the context of this conference.

### (1) Parallels with the Quantitative Revolution

A useful parallel is the "quantitative revolution" of the 1960s and 1970s. Initial research efforts were focussed on learning to use techniques developed by others, developing and/or adapting techniques particularly relevant to spatial analysis, writing computer code to ease application, and eventually, adopting generally available statistical software such as the BIOMED, SPSS, and SAS packages. While quantitative analysis struggled for acceptance initially, by 1980-85 it was both widely used and seen as essential to academic training at both the graduate and undergraduate levels. Perhaps the single-most important catalyst in this (together with empirical research that demonstrated the value of quantitative analysis) was SPSS and SAS that made statistical analysis readily accessible and included methodological instruction together with easy-to-use software.

In a similar fashion, GIS has been propelled forward by software such as ArcView, which is considerably more user-friendly than ArcInfo (at the sacrifice, at least momentarily, of analytical power). The GIS/SA interface has been enhanced by packages such as SPACESTAT and S-PLUS, but user-friendliness has been an issue. There are indications that this shortcoming is now being addressed and user-friendliness is on the upswing. However, attention to the issue appears to be a recent occurrence.

The intersection of academic courses is another issue that arises by considering quantitative revolution parallels. That involved methods courses, but also use of those methods in substantive courses and published research that provides a "demonstration effects". In academic settings today, GIS courses are generally separate from Spatial Analysis (usually statistics) courses; reference to, or use of, these in substantive courses has fallen measurably; and this is particularly true of GIS and SA together, the GIS/SA interface. In my opinion, an important force here is the limited use of GIS/SA in substantive academic research and/or confinement of its use to a small set of substantive research problems.

(2) Types of GIS Efforts

Elsewhere (September 1996 President's Column, AAG Newsletter, "The G in GIS -- Getting It Right") I postulated that GIS, from the perspective of Geography, involves three types of endeavors --

Routine-Descriptive GIS, using GIS software to make maps, diagrams, and the like;

Analytical-System Design GIS, joining GIS with statistics, cartography, information retrieval, and similar tools to answer substantive questions of a scholarly or applied nature;

Technical-System Development GIS, advancing GIS software, analytic systems, etc.

My strong feeling is that the competitive advantage of Geography lies in the second of these, the GIS/SA interface and its application to substantive research.

(3) Gains from the GIS Revolution

There have been many gains from the GIS Revolution -- NSF's initiative towards a National Center for Geographic Information Analysis (NCGIA); Ohio State's Center for Mapping; GIS as a central ingredient in government and business efforts involved with environmental management, urban planning, facility location, marketing, transportation logistics, and the like; the numerous GIS software companies and business ventures (e.g., ESRI, Business Geographics Conference, "Business Geographics" and "GIS World" publications); a geographer at the helm of the US's Census 2000.

But in my mind, two specific changes in recent years highlight the impact of the GIS Revolution. First, the visibility of Geography, as a discipline, has been tremendously enhanced. Second, maps and the portrayal of information in map form has taken on new meaning for the population at large, not to mention scientific endeavors outside of Geography.

This raises a vital question. What should Geography do, now and in the future, to get the G in GIS right so as to position ourselves for the next epoch -- to leverage and build our present advantage such that GIS gains are not simply a wave that passes?

(4) GIS/Spatial Analysis in Geography and Available Analytical Tools

My observation is that the use of GIS/SA is highly bifurcated -- (a) either reasonably sophisticated, involving technical expertise such as programming for special uses and the like, or (b) "routine map-making", e.g., choropleth maps that are rarely linked to spatial analysis beyond subjective visual impression of map patterns. The "routine map-making" exercise sometimes includes statistical analysis, but generally as a separate, rather than GIS-linked, endeavor.

Indeed, the use of even simple GIS/SA approaches would be a step towards resolving the aforementioned bifurcation. Examples of such approaches might include the use of overlay mapping, calculating the degree of correspondence between two distributions, isorithmic mapping that generalizes spatial pattern, residuals from regression, or spatially varying parameters (termed, I believe, "spatially-weighted regression" in current literature).

This is a critical observation. Even simple combinations of GIS and spatial analysis are not being exploited and/or widely used. The sense of a "missed opportunity" is very obvious.

As noted, readily accessible GIS/SA software is critical to reversing the current situation, as are teaching GIS/SA and using it in published research on substantive issues.

A related concern is sensitivity by other physical/social scientists to issues in the GIS/SA realm. In general, not widely understood or appreciated are the effects on analytical outcomes of areal size (boundaries and areal units cross-sectionally and/or in terms of their change over time; the modifiable areal unit problem), spatial resolution of data, and the meaning of contextual analysis from a spatial frame of reference (done in a manner that passes as spatial analysis for many, but would not satisfy the criteria of a GIS/SA geographer).

(5) Who Uses GIS/SA in Scholarly Research that is Substantively Motivated?

From my knowledge base, the use GIS/SA in scholarly research that is dominated by a substantive concern (and I emphasize scholarly rather than applied research, a certain brand of scholarly research at that) is more prevalent outside of geography; e.g., in anthropology, demography, epidemiology, sociology. The basis for this observation is simply scanning major geography and regional science journals of general interest relative to parallel journals of other disciplines; also from reviewing research proposals of many disciplines.

This is of critical importance. Disturbing in its own right is the point that GIS/SA might be used, and appreciated, more widely by substantively oriented professionals other than geographers-regional scientists. In addition, however, I see evidence that GIS is increasingly being disassociated from Geography; that others take it as simply another tool, and the connection to Geography per se is diminished, if not lost.

On its face, the present balance exemplifies some of the issues I've outlined earlier. There is evidence, that the balance is shifting. More centrally, the point remains that we need to carry the GIS/SA revolution forward, stimulate its wide and informed use among geographers and others, and continue to gain credit and recognition for this important tool, for the G in GIS

(6) What is the Problem, How Can It Be Remedied?

The responses to this question are embodied in earlier comments.

First, much of the substantively motivated research being done is published in specialty rather than general journals. Without neglecting the former, spreading an understanding of GIS/SA utility to a broad audience is highly important.

Second, we need readily accessible, user-friendly software that embodies the GIS/SA interface, including facilitated coupling with widely-used software such as ArcView. Moves in this direction are underway, as represented by software such as SPACESTAT and S-PLUS, mentioned earlier. It also is possible that GIS/SA laboratories will facilitate moving this interface forward, and that software companies will include greater spatial analytic properties in their product. As an example, the latest version of ArcView allows isorithmic mapping; but obviously, the GIS/SA interface must go well beyond that.

Third, we need courses, instructional direction, and workshops that focus specifically on the GIS/SA interface. As noted, our course structure tends to emphasize GIS and SA separately. Integration is needed. The one workshop of which I'm aware (University of Michigan) is focussed on social science overall, draws few geographers, but in consonance with the message here, I think it tends to draw end-users rather than methodologists per se.

Fourth, we need substantive research, carried out by non-specialists in GIS/SA (or specialists), that provide a guide, "demonstration effect", and inspiration for others. Ultimately, the success of revolution is its widespread use -- by day-to-day researchers who constitute the majority of Geography and our various disciplines. Technical research and training professionals with technical expertise remains essential, but equally essential is a base of researchers that appreciates, motivates, and uses the product of the specialists' work.

**Aileen Buckley**
Department of Geography, University of Oregon, Eugene

## Focus on Field-Based Geographic Analysis

Modern geographic analysis has evolved to incorporate computerized techniques for the collection, management, and processing of spatial data; however, increased focus on digital technology has caused remote computation to sometimes replace rather than augment more traditional in-situ field studies. More and more environmental research is being conducted in a virtual environment rather than the natural environment. This is lamentable given the long and distinguished history of field-based geographic research and problem solving. Ideally, the advantages of both field study and computation would be merged in a digitally-equipped, field-based approach to geographic observation and analysis.

Advantages of studying the environment in the field include multi-sensory observation (simultaneously seeing, smelling, hearing, touching, and even tasting the environment), process observation (seeing the components and mechanics of hydrologic, geomorphic, climatological and other natural processes), pattern identification (identifying similar and dissimilar patterns across spatial and temporal scales), and integrated observation (seeing the environment and its component parts as well as their integration). Advantages of analyzing the environment digitally include virtual experimentation (tinkering with the system without creating adverse effects that could be long-lasting or irreversible), multi-scale analysis (investigating the environment from a range of spatial and temporal scales from local to regional to continental and even to global), and spatio-temporal analysis (examining change over time through simulation, modeling, or animation).

There are also disadvantages related to both types of analysis. For example, virtual observation is a generalization of reality (reduction of the richness of the natural environment is required to digitally encode the data), it is data dependent, and it is limited by technology (capabilities and availability of both hardware and software), as well as the ability to make use of the technology. Real observation is scale-limited (observation is limited to what can be seen from a select point or points in space), static (observation is limited to the point or points in time when the observations are made), and often passive (experimentation with the environment is difficult or impossible due to logistical or legal limitations).

Introduction to the integration of field studies and computerized geographic analysis with respect to environmental studies is the focus of my position paper. Optimally, environmental analysis would incorporate both approaches -- maximizing the strengths of each while minimizing the limitations. This integrated approach is rarely introduced in academic settings, but increasingly utilized in practical situations. Geographers and other spatial analysts should be able to compare these two views of the environment, understand the strengths and limitations of each, and test and assess the value of their application.

Specific questions of interest include: What is the status of our ability to perform computational analysis while in the field? How can the environmental information collected at a site be incorporated into existing databases while retaining standards for accuracy and metadata documentation? How can the contextual information about a site be encoded for computational analysis? What kinds of geographic analyses exist or need to be developed specifically to enhance field studies?

These are but a few of the many research questions related to field-based geographic analysis. Because a colleague and I are in the process of developing a course to be taught spring term focussing on field mapping and analysis, we have identified these as our current questions of primary interest. We would very much welcome any insight into them that could be gained from the Varenius Workshop on the Status and Trends in Spatial Analysis. At the least, we would like to draw attention to the need for further development of spatial analysis in this area.

**Peter A. Burrough**
Utrecht Centre for Environment and Landscape Dynamics (UCEL),
Faculty of Geographical Sciences,
Utrecht University,
The Netherlands.

## Beyond GIS: the development of spatial analysis tools for modelling the physical environment.

During the last 10 years the University of Utrecht in the Netherlands has pursued a continuing strategy concerning the development of methods and software tools for spatial and temporal analysis. These tools have been directly linked to the needs of environmental scientists working in fields as diverse as radioecology, epidemiology, hydrology, soil science, sedimentology and physical geography. The driving forces behind the conceptual and software developments have often come from specific user needs; persons responsible for the quality of the environment, the protection of resources, or the avoidance of disasters have been increasingly turning to information technology to supply them with the data, the tools and the models to help them predict how landscapes may respond to natural or anthropogenic changes.

Standard GIS tools (by which I mean those for dealing with digital versions of paper maps and remotely sensed data) have provided useful means of data storage, data retrieval and data visualisation, together with a specific, but limited set of tools for data analysis. Today, the standard spatial entities in many environmental databases are still the supposedly the simple points, lines or homogeneous map polygon whose attributes can be analysed with many logical and mathematical tools. Many current users of digital environmental data still have few ideas of spatial interactions and spatial-temporal change, or of variation and data quality, so the methods provided by standard commercial software packages suit their needs. This situation is reinforced because computer scientists responsible for software development often view environmental data as being similar if not identical to other kinds of spatial data, such as is encountered in utility applications or land ownership systems. So long as this is the situation, there will be little reason for vendors to provide methods of analysis for which there is little demand.

At Utrecht, we realised that in order to deal with many kinds of spatial-temporal problems there had to be new developments to supplement the spatial and temporal analysis methods provided by standard GIS packages. We chose to deal specifically with the following:

1. The modelling of dynamic processes in space and time
2. Geostatistical interpolation and simulation

3. Exploratory data analysis
4. Error propagation in spatial modelling
5. Multivariate indices of spatial patterns and spatial change, in particular using methods of fuzzy logic.
6. Visualisation of spatio-temporal processes
7. Educational aspects of spatio-temporal modelling

**1. The modelling of dynamic processes in space and time**

Many environmental scientists working with dynamic processes (e.g. groundwater flow, erosion, runoff, etc.) have used GIS as a source of data which are downloaded to a model (e.g. MODFLOW). The model is run and results are returned to the GIS for display. This treats the model as a black box: it is fine if you can accept the way the model works and can supply the data it needs. On the other hand, if you want to change the model you need the source code and skills in computer programming.

Of course you can always write your own model in a language of your choice, but not everyone is a skilled programmer or has time to put together large amounts of code. So we realised that there would be many advantages to creating a generic tool for spatial-temporal modelling. Such a tool would make use of the command line interface common in raster GIS, but would provide a higher level programming language in terms most scientists could understand. A generic tool (a spatial-temporal version of MATHCAD) would:
a) Make writing and modifying models easier
b) Standardise the model interface
c) Optimize links between commands, models and the database
d) Provide a sound basis for teaching and research.

The dynamic modelling tool is called PCRASTER (http//www.geog.uu.nl/pcraster). It operates in raster mode and contains more than 150 spatial operations drawn from the rich resources of map algebra, cellular automata, hydrological routing, image filtering and so on. Additional routines can be supplied by the user via a plug and play interface. The main developments were done by Willem van Deursen (1994) and Cees Wesseling and have been continually added to by Cees and his colleagues since. PCRaster is now used by many government institutes and universities (from the European Union to individual researchers) to supplement standard GIS. It has been used in applications as diverse as the reactions of large river catchments (the Rhine, Bramaputra) to possible climatic change, nutrient flows in large catchments, soil erosion at scales from the Mediterranean to metre-square plots, the modelling of deltas and river meandering, landslides, the dispersion of plants and animals, predator-prey interactions and many more. PCRaster also enables the user to view the results of spatio-temporal models as 3-4D movie-like ‘draped’ displays so that the ways a landscape reacts to the various processes can easily be seen. PCRaster grids can be very easily exported to ARC VIEW. Information about the theory and applications of PCRaster can be found in Wesseling et al 1996, Burrough and McDonnell 1998, and Burrough 1998.

## 2. Geostatistical interpolation and simulation

Most spatial models that are run either in PCRaster or other formats require space to be discretised, either with regular cells (finite difference modelling) or defined entities or polygons (finite elements). In most cases data need to be collected from sparsely located points and then interpolated to fill the gaps. Conventional methods of interpolation are quick and dirty, and the methods of geostatistics provide a rich suite of tools for optimal interpolation of static spatial patterns. One main advantage is that geostatistical methods also give an indication of the quality of the interpolation and the errors associated with it.

Gstat, written by Edzer Pebesma, is a very comprehensive geostatistical toolkit that provides means for variogram estimation and fitting, and most commonly encountered forms of kriging interpolation including point and block estimation, simple, ordinary and universal kriging, indicator functions and stratification according to external criteria. Gstat uses the same spatial data format as PCRaster so that both the interpolated surfaces and the information about strata can be easily exchanged. Data input is via the well-known Geo-Eas format.

Gstat also includes methods for conditional simulation of spatial surfaces, which provides means for studying the role of either random or spatially coordinated errors in modelling. This provides Monte Carlo methods for following the propagation of errors in the dynamic PCRaster models. Some dynamic models with local interactions (e.g. river meandering or the modelling of alluvial fans and deltas) need a stochastic seed to get them started and Gstat provides ways of creating these randomised inputs.

More information about Gstat and how to get it can be found at gstat-info@geog.uu.nl

## 3. Exploratory data analysis

The provision of hyperlinked windows in statistical packages and in programs like ARC VIEW have greatly simplified the detection of errors in data and the amount of insights a user can get before carrying out complex analyses or modelling. At Utrecht we did a certain amount of work with John Haslett's group in Trinity College Dublin on the addition of geostatistical analyses to his REGARD programme (Gunnink and Burrough 1996). Unfortunately the Macintosh software could not easily be transferred to Windows so developments in REGARD ceased. Today, programmes like Yves Pannatier's VARIOWIN (not a Utrecht product) and S-plus, SPSS, etc. provide much useful exploratory data analysis tools that are easy to use in conjunction with PCRaster and Gstat.

**4. Error propagation in spatial modelling**

Gerard Heuvelink's work on error propagation in spatial modelling (now published by Taylor and Francis, Heuvelink 1998) was a pioneering attempt to link Geostatistics and GIS in such a way that one could identify the different sources of uncertainty in the results of GIS models, and the magnitude of the contributions from each source. By linking this work to previous work on the optimisation of sampling networks by McBratney and Webster (1981) it is possible to carry out a cost-benefit analysis of different combinations of interpolation methods and data configurations (See Burrough and McDonnell 1998, Chapter 10).

**5. Multivariate indices of spatial patterns and spatial change, in particular using methods of fuzzy logic.**

Another major line of research at Utrecht has concerned the applications of multivariate methods for classifying spatial patterns. While most statistical packages provide factor analyses and numerical clustering tools, few provide methods for fuzzy classification. Imposing the rules of an existing fuzzy classification on mapped data is little more than a standard computational operation in GIS, but deriving an optimal, overlapping fuzzy classification from point data requires other means. In common with many other researchers in this field, we have used the methods of fuzzy k-means (see Burrough and McDonnell 1998, Chapter 11). We have demonstrated that in order to achieve coherent patterns of multivariate groups it is not only essential to have a good clustering in data space, but also a strong spatial correlation structure (as expressed by the variogram).

Fuzzy membership values computed from point data can of course be interpolated by geostatistics to space filling grids, and the interpolation imposes a certain degree of spatial continuity. We have used this method for multivariate classification of soil, geochemical data and crop yield variations (multiyear results of different crops with harvesting aided by GPS). When data are taken from continuous surfaces such as DEMs, the inputs are already spatially correlated and variogram analysis is not necessary. Recently, together with Bob MacMillan and John Wilson, Pauline van Gaans and myself have used PCRaster (for the derivation of other DEM attributes like slope, plan and profile convexity, ridge proximity, together with simulation modelling of derived drainage networks and fuzzy k-means to create stable classifications of landforms. Applications range from a 150 ha site in Alberta, Canada to a 8000km2 plus area of the Yellowstone National Park. Independent tests of the classification demonstrate stability with extension to neighbouring areas, and also in terms of ecological properties of the derived units.

## 6. Visualisation of spatio-temporal processes and patterns

Having clever methods of modelling and analysis is no good unless you have the appropriate tools to present the results to the user. Cees Wesseling and Victor Jetten have developed a range of display tools, based on games routines, for the dynamic display of spatial-temporal models.

## 7. Educational aspects of spatio-temporal modelling

We have been teaching GIS and Geostatistics to both undergraduates and graduates for many years. Since 1996 we have also been teaching the methods of dynamic modelling to external and internal researchers and PhD candidates, and since 1997 to final year students so that they can use these methods in their major fieldwork studies. The results have been very encouraging indeed. An ongoing project is to create a series of virtual landscape tools in which students can explore how landscapes may react to short term and long term changes in control parameters. This enables them to follow processes that may take place very quickly (e.g. raindrop splash) or processes that take millions of years (e.g. tectonic uplift). By combining PCRaster models with Gstat functionality in a user-friendly shell (currently Powerpoint) it is easy and effective to link models to explanatory text, figures, photos and video. Another ongoing project being led by Derk Jan Karssenberg and Cees Wesseling is our contribution to an EU-sponsored Distance Learning project, which is being coordinated by Joao Ribeiro da Costa at the New Technical University in Lisbon, Portugal.

## 8. New developments.

New developments include the provision of a complete Windows interface, a 3D-4D database structure to deal with issues arising in sedimentology and erosion, improved multivariate methods, better visualisation tools, etc.

## References.

Burrough, P.A. and A. Frank (Eds.) Geographic Objects with Indeterminate Boundaries. Taylor and Francis 1996.
Gunnink J. & Burrough, P.A. 1997. Interactive spatial analysis of soil attribute patterns using Exploratory Data Analysis (EDA) and GIS. In: M. Fischer, H.J. Scholten and D. Unwin. (eds). Spatial Analysis in GIS.. Taylor and Francis (1996). 87-100.
Burrough, P.A.. Environmental Modelling with Geographic Information Systems. In: Innovations in GIS 4. Editor Zarine Kemp, Taylor and Francis 1997 , pp 143-153.
Burrough, P.A. and J. Swindell. 1997. Optimal mapping of site-specific multivariate soil properties. In: J.Lake, G. Bock and J. Goode (Eds) Precision Agriculture: spatial and temporal variability of Environmental Quality, Proc: CIBA Foundation Symposium 210,

John Wiley and Sons, Chichester, 1997. 208-220.
Burrough, P.A. & R.A. McDonnell. Principles of Geographical Information Systems (2nd edition) Oxford (February 1998), 333p., reprinted August 1998.
Burrough, P.A.. Dynamic Modelling and GIS, Chapter 9, In: P.Longley et al.(Eds) Geocomputation: a Primier. Wiley 1998, pp165-192
Burrough, P.A., P.van Gaans & R. Hootsmans. Continuous Classification in Soil Survey: Spatial Correla-tion, Confusi-on and Boundaries Geoderma 77: 115-35. 1997.
Burrough, P.A., P.F.M. van Gaans and R.A. MacMillan. High-resolution landform classification using fuzzy k-means (Journal of Fuzzy Sets - in press 1998/9:
Burrough, P.A., P.F.M. van Gaans, J. Wilson and A.J. Hansen. Fuzzy k-means classification of digital elevation models as an aid to forest mapping in the Greater Yellowstone Area, USA (in preparation)
Pebesma, E. and Wesseling, C.G. 1998. Gstat: a program for geostatistical modelling, prediction and simulation. Computers and Geosciences 24: 17-31.
Heuvelink, G.B.M. 1998. Error Propagation in Environmental Modelling with GIS. Taylor and Francis
Van Deursen, W.P.A. 1994. Geographical Information Systems and Dynamic Models. Netherlands Geographical Studies No. 190.

**Dianne Cook**
Department of Statistics, Iowa State University, Ames

## Visualizing Multivariate Spatial Data

There are numerous software tools that provide excellent computing systems for different aspects of spatial analysis. Taking advantage of existing expertise allows us to concentrate on developing methods and software for the unique aspects of spatially referenced data. We will describe the components where we believe there is mature expertise, and how we make use of these to provide an integrated analysis environment.

The geographic information system (GIS) provides an "anchor" of good database tools, and map drawing facilties, for spatially referenced data. We use this to select, or do simple manipulation of the data, and provide sophisticated maps of the spatial domain. The GIS is important also for maintaining the frame of reference to the data - a good map provides context for the data, which can otherwise get lost in statistical modeling and graphics.

A statistical analysis system (for example, S-Plus, SAS, XploRe, XLispStat, DataDesk) is used for modeling the trends and spatial dependence.

A visualization system (eg XGobi, DataDesk, XLispStat, XmdvTool, cdv) provides quick exploratory analysis and diagnostic checking for the model. The graphics need to be interactive, with several facilties for linked brushing, and dynamic to rotate the data through high-dimensional space. We should be able to quickly examine spatial dependence plots, models and residuals, as well as the multiple raw variables. We have a fairly broad variety of tools for extracting patterns in multivariate data. *There needs to be a lot more research on the types of plots that can extract multivariate spatial trends and dependence*. There are variogram clouds and cross-variogram clouds, which give information on individual and pairwise spatial dependence. But pairwise analysis of high-dimensional data is inadequate, so we suspect we will find that pairwise analysis of spatial dependence will be inadequate. So we need to devise new approaches to visually exploring spatial dependence amongst several variables.

All three systems need to be seamlessly linked. The scatterplot needs to be linked to the GIS map - brushing in the scatterplot should instantaneously update the map view and brushing in the map view should instantaneously update the scatterplot. We should be able to display the model overlaid on the geographic domain, and toggle between the model and residual surface. We should be able to calculate local statistics and make plots of these linked to the map.

**Noel Cressie**
Department of Statistics
The Ohio State University

## Mapping Rates Associated with Polygons

Suppose the geographic data under investigation are rates associated with polygons. For example, disease incidence, mortality, unemployment, per capita income, and census undercount data are rates. In these examples, polygons are typically states or counties, which are political entities that usually have nothing to do with the etiology of the phenomenon under study. Because the bases of the rates vary with polygon, one is faced with a comparison of data whose variabilities are highly different. For example, an unusually high or low rate for a polygon may be due to very few base counts. Thus, when smoothing reported rates to predict true rates, it is important to take into account this geographic heteroskedasticity.

Spatial analysis of problems of this sort can be handled very naturally through hierarchical statistical modeling, where there is a measurement process at the first level, an explanatory process at the second level, and a prior process at the third level. The resulting models are heteroskedastic and spatial, and the method of statistical analysis is Bayesian.

In our paper, we shall feature epidemiological data, reflecting the importance of disease mapping to society in general. Here, the "polygons" are known as "small areas", which has come to mean any group of regions whose whole makes up a larger region of interest. There are a number of issues related to the display, analysis, and interpretation of spatial epidemiological data that we believe are important:

* Improved small-area estimation with a focus on identifying extreme values.

Maps constructed using raw disease-incidence rates or standardized rates do not account for variation in the precision with which these rates estimate true underlying rates, because there are unequal numbers of person-years-at-risk across small areas. Hierarchical statistical analysis avoids this problem because the resulting estimates average small-area disease-incidence rates with regional or national data. Unfortunately, the resulting estimates for low-population areas can be "overly smooth" in the sense that they are less likely to be identified as locations of increased risk when they do in fact have high risk. Through the use of appropriate loss functions, we propose to construct small area estimates that facilitate the identification of areas of high risk.

* Assessing the fit of the statistical model and determining if high-risk locations have unusually high risk.

The use of statistical models to provide improved small-area estimates introduces the chance that model misspecification will lead to misleading or erroneous policy conclusions. Specifically, it remains to determine whether regions identified as having high risk for disease incidence indicate model failures or new potential risk factors.

* Relating small-area data to point-level epidemiological mechanisms.

Disease incidence or mortality data is usually reported for small areas, although the increased emphasis on data collection makes it likely that individual incidence data will be available in the future. Regardless, data regarding environmental risk factors are likely to be collected on different geographical scales than the disease-incidence data. Moreover, certain risk factors are determined at the individual or point level.

Statistical models are needed that allow for the integration of individual mechanisms with small-area data and for the possibility of aggregation of some small areas into larger small areas. GIS will play an important role in managing data of different aggregations and in displaying the results of the hierarchical statistical analyses referred to above.

Research presented in this talk is joint with Hal Stern and Deanne Reber of the Department of Statistics, Iowa State University.

**A. Stewart Fotheringham**
University of Newcastle

## Stressing the Local

## 1. Introduction

Although my research interests have meandered between spatial interaction/spatial choice modelling and spatial statistics, a common theme to much of my work has been an interest in identifying and understanding *differences* across space rather than *similarities*. The development of spatial disaggregations of global statistics was the subject of my PhD on origin-specific distance-decay parameters in spatial interaction models (still a prime interest) and it is the subject of recent research at the University of Newcastle on *Geographically Weighted Regression*. The concern for the local encompasses the dissection of global statistics into their local constituents; the concentration on local exceptions rather than the search for global regularities; and the production of local or mappable statistics rather than on 'whole-map' values. This trend is important not only because it brings issues of space to the fore in analytical methods, but also because it refutes the rather naive criticism that quantitative geography is unduly concerned with the search for global generalities and 'laws'.

Obviously, local forms of spatial analysis are important to GIS because they result in geocoded output that can be mapped. It could also be claimed that some of the impetus for the development of local statistics derives from the growing interest in integrating advanced forms of spatial analysis and GIS (Fotheringham and Charlton, 1994; Fotheringham, 1994; Fotheringham and Rogerson, 1993).

The theme of much of my research has been that when analysing spatial data, it may be incorrect to assume that the results obtained from the whole data set apply equally to all parts of the study area. Interesting insights might be obtained from investigating spatial variations in the results. Simply reporting one 'average' set of results and ignoring any possible spatial variations in those results is as limiting as reporting a mean value of a spatial distribution without seeing a map of the data

## 2. The Nature of Local Variations in Relationships

There are at least three reasons to question the assumption of stationarity in spatial data analysis and to allow variations in observed relationships, as measured for example by parameter estimates in a regression model. The first and simplest is that there will inevitably be spatial variations in observed relationships caused by random sampling variations. The contribution of this source of spatial non-stationarity is not usually of great interest but it does need to be accounted for by significance testing. That is, we are only interested in relatively large variations in parameter estimates which are unlikely to be caused by random sampling and which therefore constitute interesting spatial non-stationarity.

The second is that, for whatever reasons, some relationships are intrinsically different across space. Perhaps, for example, there are spatial variations in people's attitudes or preferences or there are different administrative, political or other contextual issues that produce different responses to the same stimuli over space. This idea that human behaviour can vary intrinsically over space is consistent with post-modernist beliefs on the importance of place and locality as frames for understanding such behaviour.

The third reason why relationships might exhibit spatial non-stationarity is that the model from which the relationships are measured is a gross misspecification of reality and that one or more relevant variables are either omitted from the model or are represented by an incorrect functional form. This view, rather more in the positivist school of thought, runs counter to that discussed above in that it assumes a global statement of behaviour can be made but that the structure of our model is not sufficiently well-formed to allow us to make it. In this case mapping local statistics is useful in order to understand the nature of the model misspecification more clearly. For what parts of the study region does the model replicate observed data less accurately and does the spatial distribution of these parts suggest the addition of an extra explanatory variable to the model?

## 3. Attempts to Measure Local Variations in Relationships

Within the last several years, there has been a relatively flurry of academic work reflecting the calls of Fotheringham and Rogerson (1993), Fotheringham (1992) and Openshaw (1993) for greater attention to be given to local or mappable statistics. Four areas are now described where progress has been made in this direction

### 3.1 Local Point Pattern Analysis

The analysis of spatial point patterns has long been an important concern in geographical enquiry but until relatively recently most applications of spatial point pattern analysis involved the calculation of some global statistic that described the whole point pattern and from which a conclusion was reached related to the clustered, dispersed or random nature of the whole pattern. Clearly, such an analysis is potentially flawed in that interesting spatial variations in the point pattern are subsumed in the calculation of the average or global statistic. In many instances, particularly in the study of disease, such an approach would appear to be contrary to the purpose of the study, namely to identify any interesting local clusters.

Amongst the first example of a local point pattern analysis technique was the Geographical Analysis Machine (GAM) developed by Openshaw et al (1987) and updated by Fotheringham and Zhan (1996). The basic idea is very simple and serves to demonstrate the interest in the local quite well. Within the study region containing a spatial point pattern, randomly select a location and then randomly select a radius of a circle to be centred at that location. Within this random circle count the number of points and compare this observed value with an expected value based on an assumption about the process generating the point pattern (usually that it is random). Ideally, the population-at-risk should be used as a basis for generating the expected value, as shown in Fotheringham and Zhan (1996) who use a Poisson probability model with the observed mean and the population-at-risk within each circle. Once the statistical significance of the observed count within a circle has been examined the circle is drawn on a map of the region if it contains a statistically significant cluster of points. The process is repeated many times until a map is produced containing a set of circles centred on parts of the region where interesting clusters of points appear to be located.

### 3.2 The Local Measurement of Univariate Spatial Relationships

Much of the work undertaken in exploratory graphical analysis is essentially concerned with identifying local exceptions to general trends in either data or relationships. Hence, techniques such as linked windows and brushing allow data to be examined interactively so that points appearing as outliers in various statistical displays can be located on a map automatically. Usually this type of graphical interrogation takes place with univariate distributions so that histograms or box-and-whisker displays form the basis of the graphics although scatterplots can also be linked to a map display and even 3-D spin plots can be used. No matter which exploratory technique is used, however, the aim of the analysis is generally to identify unusual data points and the focus is on the exceptions rather than the general trend. More formally, local versions of global univariate statistics have recently been developed by Getis and Ord (1992), Ord and Getis (1995) and by Anselin (1995).

### 3.3 The Local Measurement of Multivariate Spatial Relationships

The increasing availability of large and complex spatial datasets has led to a greater awareness that the univariate statistical methods described above are of limited application and that there is a need to understand local variations in more complex relationships. In response to this recognition, several attempts have been made to produce localised versions of traditionally global multivariate techniques, with the greatest challenge being to produce local versions of regression analysis.

Perhaps the best-known attempt to do this is the expansion method (Casetti, 1972; Jones and Casetti, 1992) which attempts to measure parameter ‘drift’. In this framework, parameters of a global model are expanded in terms of other attributes. If the parameters of the regression model are made functions of geographic space, trends in parameter estimates over space can then be measured (Fotheringham and Pitts, 1995; Eldridge and Jones 1991). Whilst this is a useful and easily applicable framework in which improved models can be developed, it is essentially a trend-fitting exercise in which complex patterns of parameter estimates will be missed. The output from spatial variants of the expansion method is thus a second-order set of relationships when what is required is information on the first-order relationships.

More recently, Geographically Weighted Regression (GWR) (Brunsdon et al. 1996; 1998; Fotheringham et al 1996; 1998) has been developed to extend the traditional regression framework by allowing local rather than global parameters to be estimated. That is, the model to be estimated has the general form:

$$y_i = a_{i0} + \sum_k a_{ik} k_{ik} + \varepsilon_i$$

where y represents the dependent variable, $x_k$ represents the kth independent variable, $\varepsilon$ represents an error term and $a_{ik}$ is the value of the kth parameter at location i. In the calibration of this model it is assumed that observed data near to point i have more influence in the estimation of the $a_{ik}$ s than do data located farther from point i. In essence, the equation measures the relationships inherent in the model around each point i. To calibrate the model, a modified weighted least squares approach is taken so that the data are weighted according to their proximity to point i. Thus the weighting of any point is not constant but varies with i. Data from observations closer to i are weighted more heavily than those from farther away. Hence the estimator for the parameters in GWR is:

$$\hat{a}_r = \left(x^t w_i \cdot x\right)^{-1} x^t w_i \cdot y$$

where $w_i$ is an n by n matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data for point i. It should be noted that as well as producing localised parameter estimates, the GWR technique described above will produce localised versions of all standard regression diagnostics including goodness-of-fit measures such as r-squared. The latter can be particularly informative in understanding the application of the model being calibrated and in exploring the possibility of adding additional explanatory variables to the model. A list of recent GWR publications and code is provided at the GWR web site: www.ncl.ac.uk/ngeog/nmec/GWR

### 3.4 The Mathematical Modelling of Flows

Perhaps one of the earliest, yet still misunderstood, examples of providing local information on relationships rather than simply reporting global results is the spatial disaggregation of spatial flow or spatial interaction models (Fotheringham and O'Kelly, 1989). The reason for calibrating spatial flow models is to obtain information, via the estimated parameters of the models, on how individuals make choices amongst spatial alternatives. By far the most important attribute in many spatial choice contexts is the spatial separation between the individual and the alternative - individuals are less likely to choose an alternative that is farther way, *ceteris paribus* - and an important aspect of the calibration of spatial choice/spatial interaction models is to obtain information on the rate of this 'distance-decay'. It was recognised quite early in the spatial interaction modelling literature that localised distance-decay parameters would yield more useful information on the spatial choice process than simply estimating a global interaction model (see Fotheringham, 1981 for a review). From an accumulation of empirical examples of origin-specific parameter estimates, it has proven possible to map trends in parameter estimates that have led to the identification of a severe misspecification bias in the general spatial interaction modelling formula (Fotheringham, 1984; 1986). It is worth stressing that such misspecification only came to light through an investigation of spatial variations in localised parameters and would have been missed in the calibration of a global model.

## 4. Summary

The relative explosion of attention to the ‘local’ rather than the ‘global’ in quantitative geography is interesting for several reasons. It belies the criticism that the quantitative approach is only concerned with the search for broad generalisations and not with identifying local exceptions. It links quantitative geography with the powerful visual display environments of various GIS and statistical graphics packages where the all-important display is the map. It also allows quantitative geographers to explore relationships in different ways as a guide to a better understanding of spatial processes and finally it affords the exciting opportunity of developing new statistical approaches to spatial data analysis

## 5. References

Anselin, L. 1995: Local Indicators of Spatial Association - LISA Geographical Analysis 27, 93-115.

Brunsdon, C.F., Fotheringham A.S. and Charlton, M.E. 1996: Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity, Geographical Analysis, 28: 281-298

Brunsdon C.F., Fotheringham A.S. and Charlton, M.E. 1998 Spatial nonstationarity and autoregressive models Environment and Planning A 30: 957-973

Casetti, E. 1972: Generating Models by the Expansion Method: Applications to Geographic Research, Geographical Analysis 4, 81-91.

Cleveland, W. S. 1979: Robust Locally Weighted Regression and Smoothing Scatterplots Journal of the American Statistical Association 74, 829-836.

Eldridge, J.D. and J.P. Jones III 1991: Warped Space: a Geography of Distance Decay Professional Geographer 43, 500-511.
Fotheringham, A.S. 1981: Spatial Structure and Distance-Decay Parameters, Annals of the Association of American Geographers, 71, 425-436.

Fotheringham, A.S. 1984: Spatial Flows and Spatial Patterns, Environment and Planning A, 16, 529-543.

Fotheringham, A.S. 1986: Modelling Hierarchical Destination Choice, Environment and Planning A, 18, 401-418.

Fotheringham, A.S. 1992: Exploratory Spatial Data Analysis and GIS, Environment and Planning A, 24, 1675-1678.

Fotheringham, A.S. 1994: On the Future of Spatial Analysis: The Role of GIS Environment and Planning A Anniversary Issue: 30-34.

Fotheringham, A.S. and Charlton, M.E. 1994: GIS and Exploratory Spatial Data Analysis: An Overview of Some Research Issues, Geographical Systems, 1, 315-327.

Fotheringham, A.S. and O'Kelly, M.E. 1989: Spatial Interaction Models: Formulations and Applications London: Kluwer.

Fotheringham, A.S. and Pitts, T.C. 1995: Directional Variation in Distance-Decay Environment and Planning A 27, 715-729.

Fotheringham, A.S. and Rogerson, P.A. 1993: GIS and Spatial Analytical Problems. International Journal of Geographic Information Systems 7, 3-19.

Fotheringham, A.S. and Zhan, F. 1996: A Comparison of Three Exploratory Methods for Cluster Detection in Spatial Point Patterns Geographical Analysis in press.

Fotheringham, A.S., Brunsdon, F.C. and Charlton, M.E. 1997: Measuring Spatial Variations in Relationships with Geographically Weighted Regression, Chapter 4 pp 60-82 in Recent Developments in Spatial Analysis, Spatial Statistics, Behavioural Modelling and Computational Intelligence eds. M.M. Fischer and A. Getis, Berlin: Springer-Verlag.

Fotheringham, A.S., Brunsdon, C.F., and Charlton, M.E. 1999 Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis, Environment and Planning A 30: 1905-1927.

Getis, A. and J.K. Ord, 1992: The Analysis of Spatial Association by Use of Distance Statistics, Geographical Analysis 24, 189- 206.

Jones, J.P. and Casetti, E. 1992: Applications of the Expansion Method London: Routledge.

Openshaw, S. 1993: Exploratory Space-Time-Attribute Pattern Analysers, in Spatial Analysis and GIS , eds. Fotheringham A.S. and Rogerson, P.A, 147-163, London: Taylor and Francis.

Openshaw, S., Charlton, M.E., Wymer, C. and Craft, A.W. 1987: A Mark I Geographical Analysis Machine for the Automated Analysis of Point data Sets, International Journal of Geographical Information Systems 1, 359-377.

Ord, J.K. and A. Getis, 1995: Local Spatial Autocorrelation Statistics: Distributional Issues and an Application, Geographical Analysis 27, 286-306

**Andrew U. Frank**
Geoinformation, Technical University of Vienna, Vienna, Austria

## GIS FOR POLITICS: Requirements for GIS in Political Decision Making.

GIS are excellent tools to collect spatial data, for example about the environment, but they lack some of the capabilities required to access and analyze the data and produce the information required by the political process for policy making. Systems like MARS in Europe and comparable systems in other countries produce regular analysis of the actual situation of some sector; e.g., agriculture or the land-use, but they are limited to this sector and to a post-fact analysis. They are not well suited to cross data coming from different sources (e.g., remote sensing and administrative data sources) and to answer to information needs of other sectors, even less to extrapolate and to predict future situations. For GIS to gain more attention in the public debate, they must contribute information, which can help to answer the most pressing political question, e.g., the reorganization of the European agricultural market or efforts to stop the degradation of the natural environment. What policy makers need is a tool, which helps them to assess the effects of actions considered and to evaluate different plans. GIS provide the background for such analysis, but they must be extended to include geographic facts in a spatio-temporal context and allow what-if questions, which in turn require functional models and simulation tools. In this paper, we analyze why current GIS are limited to spatial static facts and link this impediment to the mathematical-logical foundation of current GIS. The paper concludes with a list of formal tools, which can be used to build the future dynamic, temporal GIS which model geographic facts and processes.

**Keywords**: Environment, Interoperability, Functional Models, Simulation, Dynamic GIS, Temporal GIS

### Introduction

GIS are tools to collect and spatially integrate data. They bring together the results of various observations, often based on remote sensing, and promote the integration of data from different sources. The results are detailed descriptions of the current situation of the world.

We observe today that the use of GIS is increasing rapidly- in some countries more than 10% per year -, but not as rapidly as the enormous potential of GIS for administration, policy making and science promises. It was estimated that 80% of human decisions contain a spatial component [2] , and thus in most decisions a GIS could contribute to improve the decision or to reduce insecurity. GIS is- despite the rapid growth- seldom used and many areas of decision-making are still going without the benefit of spatial information. This document compares the capabilities of today’s commercial GIS and the concepts used for GIS today and compares them with the requirements of administration, science and policy making. The analysis identifies the impediments of today's GIS concepts for widespread use in political decision-making.

The discussion here addresses first very general issues, and situates them within the complex of policy making, especially the agricultural and environmental policies, which are of great practical interest today in Europe, but also in other parts of the world. The impediments, which are discussed, seem to be related to the integration of data, but beyond this integration looms large the static nature of today's GIS. Indeed we will show in the conclusion that the problem of semantic integration and modeling of process are closely related.

Current commercial GIS products are optimized for the management of *static* data about the spatial world and contain increasingly spatial analysis tools. It is possible to build systems which allow the comparison of time series of spatial data[11] , but complete integration of temporal and spatial data has not yet been achieved and research efforts are underway (Chorochronos Project, see http://www.dbnet.ece.ntua.gr/~choros/). The collection of spatial data from different sources is (this has been a well-known problem for years) hindered by differences in the format under which data are stored. Data transfer technology [20] and more recent, interoperable systems [6, 17] are to redress this impediment. At the current time, the transfer of data from one system to another is usually not a problem anymore, but the integration of data with different spatial reference system, different level of detail (often described as different scale) and different collection methods is still a confusing issue. National Geographic Infrastructure efforts [7] are focusing on these issues at various levels.

The GIS we have today can answer questions what is in the world; it is the question we ask often when we have to solve mundane problems: where to find a gas station, how to drive to our holiday resort, where to locate a new grocery store, etc. For scientific analysis, we ask questions like "where are areas used for olive growing", "where are areas with high populations of pigs". These are static questions.

More engaging are questions, which point to an explanation, questions about future states and, politically most important, questions which link possible actions to future effects. "Do olive growing areas in Spain expand?", "Where are areas where pig populations decrease?" and then the ubiquitous question: "Why?" None of these questions current GIS address well.

The quality of the environment is a high political priority. Environmental GIS are built [15], but today's GIS can at best give a comprehensive picture of the current situation. The public discussion centers around the economic effects of different alternative actions and is often shocked by the cost of corrective actions to recover from environmental sins of the past (several drastic examples are now discovered in the former Soviet Union, see various contributions in the GIS 98 Brno Proceedings[16] ). The economic and social cost of environmental degradation is much higher and integrated models are required to assess such cost. Environmental degradation (in a wider sense) is often responsible for the increase in natural hazards. The politicians need to understand the consequences of their decisions and the likely cost to repair the damage of landslides, flooding, and forest fires provoked by thoughtless land cover changes. GIS should provide tools to demonstrate to politicians the effects of agricultural policies, e.g., the resulting increase in fertilizer usage if production is increased, and the secondary effects on the environment, e.g., the elevated nitrate levels in water supplies.

Some GIS are used to collect data about the natural world and its physical properties. Other GIS collect data about the social environment- many administrative systems contain a wealth of social data which can be explored. The integration of these two realms is necessary but hindered by the different traditions of natural and social sciences, which go back to fundamental philosophical differences.

In order to direct future research we have to understand the requirements of society at large. In order for GIS to gain more attention and more resources for its development, we have to show to the public that GIS can contribute to the pressing question of society. Only then, GIS can participate of the limelight of public discussion.

## GIS Contribution to Politics

What can a GIS contribute to today’s eminent political problems? Only if it can contribute to the pressing issues of the day, we can expect that GIS finds the attention of the public and the resources needed for research, data collection and management of GIS and the related technologies.

The question is therefore to analyze the relationship between political action and the interest of politicians with the contributions a GIS can or could make. Politicians are not as irrational and shortsighted as scientists like to caricature them. They have difficult decisions before them, affecting the lives of people in many different ways. They are generally interested in information to support their case. It is our obligation to provide this information in a rational, scientific way.

For this paper, we select an issue of current eminent interest in Europe as a case study. Only if the GIS can contribute to such a case, we can benefit from the public attention, which this issue currently has. In the public TV debate so far, often during prime-time news, I have never heard mentioning the contribution a GIS based study could make. I think it behooves us to analyze the reasons why.

## Case Study: European Agricultural Policy

The single issue of major importance for European politics in the next decade selected here as a case study is the reorganization of the agricultural policy (as part of Agenda 2000). Agriculture is a major part of economy for many areas of Europe, and agricultural policy affects therefore the social situation in large parts of Europe. Contribution to the market for agricultural goods amounts to one third of the budget of the European Commission. But agriculture is also identified as a major contribution to environmental pollution and agricultural policy therefore affects the state of the natural environment. Thus in agricultural policy, the interaction with the environment and the social situation must be considered- the environmental questions are linked (through agricultural processes, but this is only one particular example) linked to the social questions. This is a perfect example for the complex environmental questions we have to address in the future and I will concentrate on how GIS can be improved to contribute to solving such problems.

**Political Questions**

A rational politician confronted with a decision will consider the alternative actions and the effects these actions may have. The outcomes are assessed with respect to his constituency- the people that have elected him- and to his party associates, relevant industry, etc.

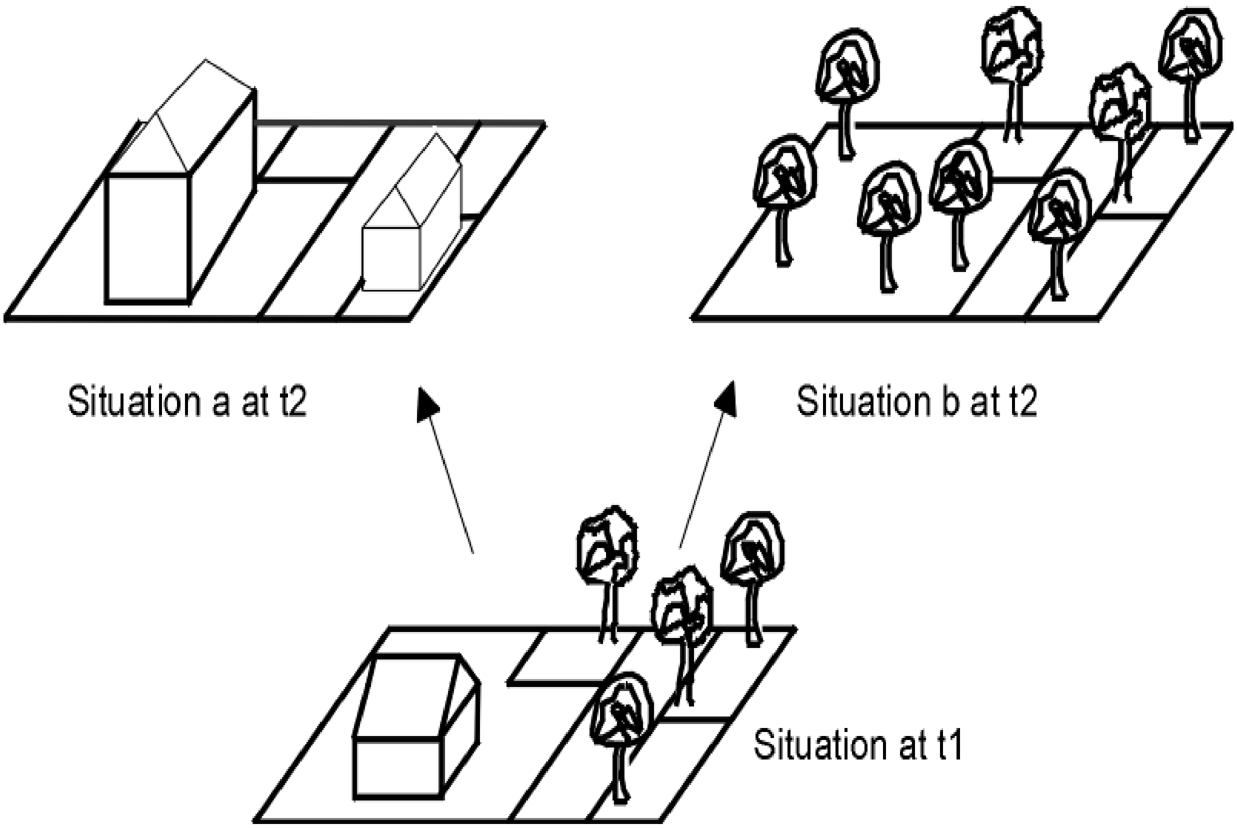


Figure 1: Two different future landscapes at t2, modeled from current situation at t1. GIS visualization allows politicians to evaluate the different possibilities.

The political debate should be about the evaluation of the outcomes for different people with different needs and the political decision should lead to an optimal set of actions which produce maximum benefits (Figure 1). Political debate is, in principle, about the evaluation of the future states; different parties may differ in their evaluation functions. Practical politics is much about strategies to achieve an optimum for one's constituency to be reelected. Mathematical game theory [22] provides a mathematical framework for the analysis of political behavior and strategies (for a more detailed analysis [23] ).

Current political debate is not only about the differences in the evaluation of outcomes, but often compound with (1) a debate about the description of the current states and (2) a debate about the likely effects actions have on the current state. This makes political debate more complex and confusing than necessary.

Optimally, science should provide the politicians with an agreed upon description of the facts to reduce the political confusion, and acceptable models to link potential actions to likely outcomes. Then politicians could concentrate on their primary function, namely the political evaluation of outcomes and the selection of the most beneficial one. Unfortunately, GIS cannot provide this today because we do not know (1) how to integrate data to build acceptable databases of facts and (2) GIS do not contain process models, which allow "what-if" questions to explore the outcomes of potential decisions.

## Scientific Questions

Surprisingly, scientists ask very similar questions as politicians. In theory, scientists posit a hypothesis and then formulate an experiment, which either confirms the hypothesis or falsifies it [24]. For scientific work to progress, a detailed description of the state before and after an action is required. The action is then modeled and the predicted outcome of the model compared with the observed outcome. The comparison leads to the acceptance of the model or its rejection.

## Interesting Questions to Politicians and Scientists

For the questions politicians ask they need descriptions of states, and models which link current state and actions to future states. This problem can be captured in formulae to construct a framework for the discussion of particular applications. The theory of modeling dynamic systems [26] gives a framework which is often used for the discussion of economic or global environmental scenarios, but seldom applied to environmental issues in a localized (spatially disaggregated) form. Considering a current state $s$ and actions $a$ which could be carried out. If actions are taken, then the current state is transformed by the function $f$ (which represents the complex system, e.g., the environmental, agricultural and social interactions) to become state *s* (Formula 1). Politicians apply a valuation function $ps_i$ to a state to gain an assessment of the desirability of this state (here described by a value *v*; Formula 2). The valuation function $ps_i$ can be applied to the future state *s*, which would be achieved after action *a* are applied (Formula 3). The politician's goal must be to select a set of actions, which maximize the outcome for his constituency (leaving the details of the optimality definition to the discussion in the economic theory).

$f (s, a) = s$ (1)

$v = ps_i (s)$ (2)

$v_i = ps_i (s_i) = ps_i (f (s, a_i))$ (3)

select $a_i$ for which $v_i$ is max!

The scientists' interpretation in the terms used above is: A hypothesis is a function *f*, which links current states *s* and the experiment (an action) *a*with observable outcomes *s*. If the observed outcome $\overline{s'}$ is the same as the predicted outcome *s* then the hypothesis is confirmed, otherwise it is rejected; a single experiment is sufficient to falsify a theory, but all experiments in the world cannot prove a theory and therefore Popper stresses the importance of falsification for the development of theories.

$s' = f(a, s)$ - for the predicted outcome

$\overline{s'} = \overline{f'}\ (a, s)$ - is the observed outcome after action *a* applied in state *s*

Very often, scientific theories *f* contain a number of constant parameters. Past observations are used to calibrate the parameters, such that the theory predicts optimally the past observed performance. The theory with the determined parameters can then be used as a predictive theory, technically to extrapolate into the future- assuming that the parameter values will not change.

## Commonality

We see that both science and politics are not interested in the current state of the world *per se*, but are interested in models of the world which link the current state and actions with a future state. The GIS today at best provides the description of some aspects of current state, which is useful and valuable for many applications (mostly in the administrative domain) but does not respond to the major challenges of our living environment and the conditions for human live.

Both political and scientific studies work with data with limited accuracy and in consequence accept results with limited accuracy. Scientists often use statistical tests, to differentiate between the arbitrary random effects and significant effects. Politicians request data which is *fiable* and despise data which can be easily falsified in details by their political adversaries (even if the general trend is correct): the message a politician sends must evoke confidence and if it can be shown to be false in a single detail, the trust the public places in this elected politician is lost. We lack so far an assessment of 'political fiability' which corresponds to the concept of scientific significance.

Common to both, to the scientist and the politician, is that at the end, they are satisfied with a qualitative answer: the hypothesis is confirmed or rejected; the set of actions, which leads to the optimal state, is identified. Both are interested in 'models', which link the current state and actions to future state. The terminology is not uniform, but scientists often call models theories if they are constructed in terms of a supportive science (or more detailed level of the same science).

This very general analysis provides us with a common framework for both scientific analysis and for political and administrative use of GIS. It stresses the importance of the dynamic modeling power of the GIS required.

## Requirements for Future GIS

GIS today manage data which describe the current (or some past) states of the world. In order for GIS to contribute more to the pressing problems of the world, they must be expanded. Two key requirements must be addressed:

- Integration of data, to construct the database to calibrate complex models,

- Description of process, to permit what-if questions

and we will show in this section, how these are interrelated. The next two sections will then discuss the mathematical foundations and the intellectual tools available to address these challenges. Later we will discuss the practical efforts already underway to contribute to them.

## Spatial Database Management

GIS as a special, spatial case of data management is of limited interest and limited commercial success to the software vendors. It is clearly visible that increasingly the large DBMS providers include some form of spatial data management tools in their offerings (Oracle: Spatial Data Organization; Informix: Spatial Data Blade). Spatial data management is not the core issue of GIS research any more, as it was some years ago; there is a now regular bi-ennnial conference where mostly computer scientists meet to discuss spatial access methods and related database issues of importance to GIS [28].

The research challenge today is more with the integration of spatial data into regular DBMS with minimal adaptation of the complex kernel of a full-feature, industry-strength DBMS. Proposals for solutions by Abel [1] have been tested and work well.

## Integration of Data

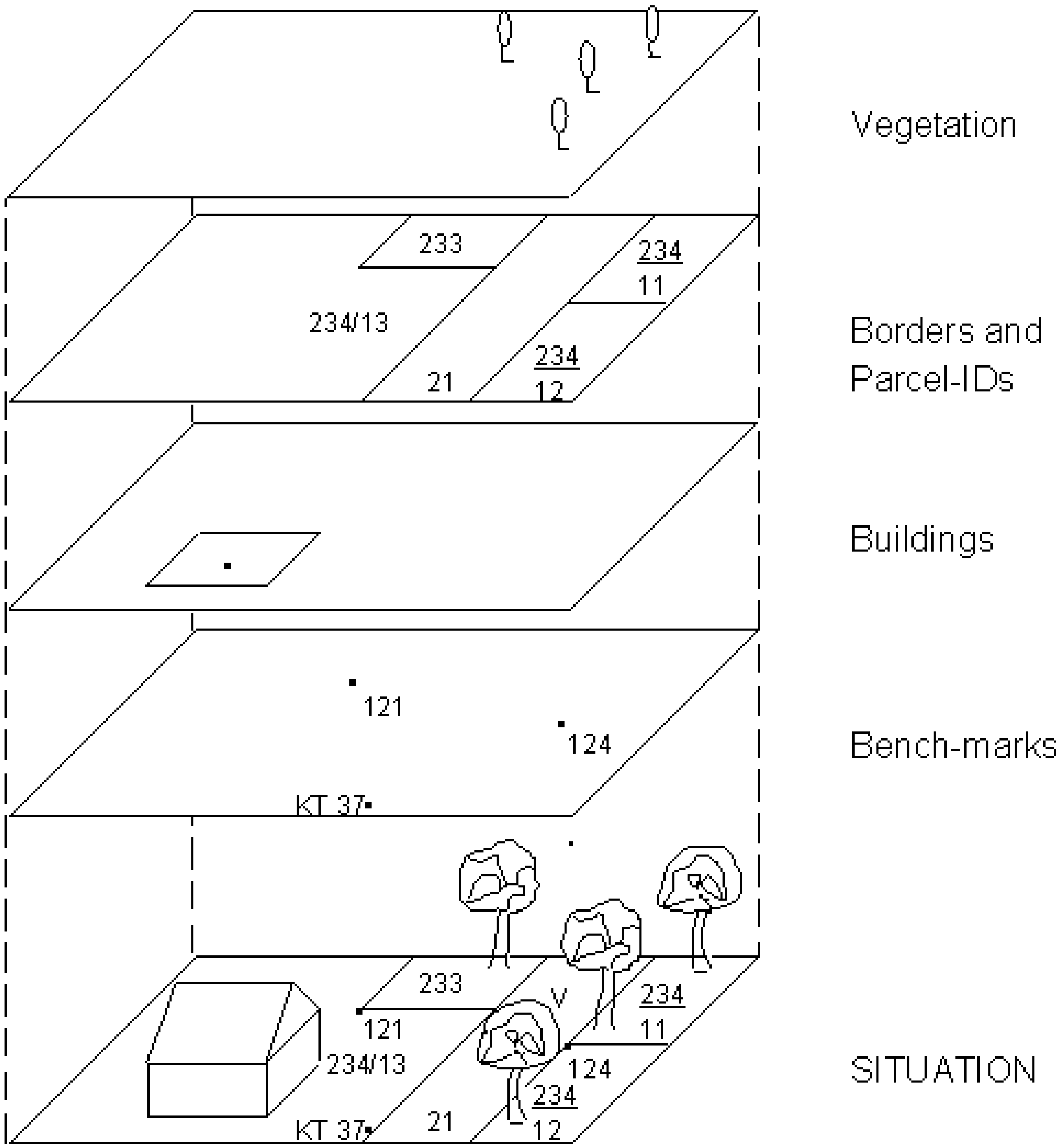


Figure 2: Different thematic representations of the same situation.

The promise of GIS is the integration of data from different sources with respect to location in space, best visualized in the often seen stack of thematic maps (Figure 2).

Practically, GIS has delivered on this promise only partially. We can integrate routinely, i.e., without major expert intervention:

- raster data with comparable resolutions and apply the operations defined by Dana Tomlin [31]
- polygonal data (coverages) if the data is of comparable spatial resolution and precision.

This is a quite limited success; it excludes most interesting data integration tasks politicians and scientists wait for. Integration works do not work automatically; if the data is of different resolution, level of detail, measurement scales etc., then the integration requires very careful analysis of the datasets by experts, is often achieved only with a very time consuming trial and error process and produces results of questionable value.

The seamless integration of raster and polygonal data is a long standing question; in principle, it can be easily resolved by translating all data to a raster format but leaves out linear data which is not easily expressed in a raster data model. The integration of data of different resolutions is theoretically equally simple: either the more detailed data set is aggregated to the level of resolution of the less detailed one or the less detailed is blown up to match the resolution of the more detailed one. The first approach invokes all the known problems of cartographic map generalization [32], which are not yet solved automatically. The second approach produces results, but their interpretation is extremely tricky.

Politicians expect us to relate spatial data from social statistics with data describing the physical environment. Political decisions are made by people for people and the models must contain the assessment of the outcome of actions for people. A common but important example is the association of population counts with the urban area- population counts are collected for administrative subdivisions, which do not coincide with the boundaries of geographic units like urban (built-up) area. Theoretically, the problem is the lack of a uniform system of reference areas. The problem is further aggravated if changes in time are considered: data collected at time t1 for the administrative units are not directly comparable to the data collected at t2 which refer to the administrative units at this time, which are often changed (see [12], in particular the paper by Jostein Ryssevik [27] ).

But even if the technical solutions are available to integrate data in a comparable format, other non-technical issues may make the integration difficult. The data may not be available for a number of administrative and legal reasons [19] and if it is available, it might be difficult to decide if it can be used. To assess the semantics and the quality of the data is again a question only experts can answer [14]. Data collected for different purposes- even if supposedly the same phenomena are observed- may differ substantially (e.g., terrain height collected by terrestrial and photogrammetrical methods; population counts from census or from daytime sampling).

### Modeling of Process

Scientists and politicians are interested in the processes which change our world, not in simple collections of facts: There are few people who read the phone book with great interest despite that it is a very extensive collection of facts! For an effective use of spatial data for science and policy making, collections of facts like the census results are important, but modeling of process is essential.

The GIS has never promised to deal with process modeling. The initial view was that GIS would manage the data and other tools would include the process models. There were concepts of systematically organized collections of methods for analytical and other purposes. Unfortunately, this concept did not materialize, because the interface between the process models and the data models were not resolved.

The impediments today are:

- GIS data is essentially static and presents a snapshot of the world;
- Lack of methods to discuss dynamic processes;
- Difficulty to link dynamic process models with the static spatial data collections.

The formal models we use today in GIS (but also in most other geographical sciences) are mostly static and do not include formal, generalized expressions of processes, which change a static situation (see discussion in next section). Simulation tools for dynamic modeling' [26], as used in other sciences (electrical engineering, economy etc.) are not widely used in geographical sciences. The lack of theory for the description of the process implies that the integration of process model and data management is complicated and no general-purpose solutions have been found yet.

## Logic Foundations to Handle Data are Static

The logical tools developed for the description of data- at least their logical structure- are extensive. There are the universally used concepts of relational databases, with the concepts of relational table, tuples and key, and the operations of selection, projection, join, union and difference, but there as old is the entity-relationship model, with entity and relations as the founding blocks.

These tools are based on and - in essence - do not go beyond their foundation in logic. In a fundamental early paper, Gallaire, Minker and Nicolas [13] gave a clean mathematical connection between databases and first order predicate calculus. It was shown that the logical framework is more powerful than relational algebra (which led eventually to practical extensions to include 'transitive closure' into the relational algebra). It made the knowledge about logical systems and the power they have applicable to the database world (for an extensive discussion of the design space for theories see [4]). The viewpoint spurred a large research effort to use Prolog or extensions of Prolog to explore data.

These efforts pointed out what cannot be modeled in a relational database. The standard logical assumptions in a RDBMS include the so-called 'closed world assumption' [25]. It says that the database is a complete picture of the world, and all things in the database are true, but also- and this is the crux for GIS- that facts which are not in the database are false. The closed world presupposes a complete knowledge of the world, which is achievable in administrative systems, but is never the case for GIS. With the closed world framework, it is not necessary to express negative facts because just leaving out a fact asserts its negation.

## Integration of Temporal Aspects in Logic Based Formalizations

Logic is an essentially static system: it describes what is implying at a given time or sometimes always. This can be extended to systems of temporal logic [33], which describes situations related to a point in time and a calculus can be formed (the situation calculus of McCarthy [18]). Formulae can then express changes from one state to the next, and logical deduction is used to connect these. The so-called frame problem, i.e., the need to express that all facts not affected by a change formula remain the same, must be addressed. With these extensions, a predicate calculus based situation calculus has a format and an expressive power, which is quite similar to the algebra based tools described next. This theoretical result has recently been verified by a M.Sc. thesis, where the dynamic characteristics of a land registration system was formalized, once using algebra and once using situation calculus [21].

## Algebra-Based Tools can Model Dynamic Situations

An algebra is a description of a set of connected operations which apply to a set of types. This is the generalized definition of algebra, introduced by Birkhoff as universal algebra. It generalizes the concept of an algebra, for example, the algebra over complex numbers with the operations +, -, *, etc. to situations where the operands of the operation have different type. We give here the well-known example of a stack, to demonstrate the principle.

A stack can accept elements pushed onto it a simple example is a stack of plates as found in any cafeteria. The operation top return the top element, the operation pop returns a stack with the top element removed (these two operations are usually merged in real computer implementations, but to achieve mathematical clarity and simplicity, they must be separated).

```
class Stack s a where -- where s a is a stack of elements a

 new :: -> s a

 push :: a -> s a -> s a
```

*pop :: s a -> s a*

*top :: s a -> a*

Algebras capture both the notions of the abstraction of objects and the abstraction of operations. The concept of an algebra is therefore fundamental when we attempt to discuss changes as operations (not just as difference between two states). For the stack, the behavior of the operations can be fully explained by a few axioms like:
*top (push (a, s)) = a - the top element after pushing an element onto the stack is the element which was pushed on.*

*pop (push (a,s)) = s - the stack returned after pushing something onto a stack and then apply a pop to the result is the same as the stack before the push operation.*

The theory of algebra can be further abstracted to category theory [3]. In category theory, the axiom
*pop (push (a,s)) = s*
would be written as
*pop. push = id,*
stating that the combination of a *push* and a *pop* operation is the identity operation (the operation, which does nothing). Both algebra and category theory provide the instruments for the description of semantics without the infinite regress to previously defined terms. They give tools which allow the definition of terms without relying on previously understood terms (which in turn, rely on previously defined terms leading to an infinite regression). Algebraic tools are therefore useful for the definition of semantics, especially across language and cultural differences in Europe.

**Object Orientation Tools**

Object Orientation is a trend in software engineering, closely related in the theory to universal algebra and category theory. Object Orientation is seen as the solution for the software crisis, and the concentration on objects and related operations is fundamental for the programming of graphical user interfaces, but also for other code. It is supported by most programming languages currently used.

UML is the new conceptual tool for the designer and a set of programs (Rational Rose) is available [10]. The concepts are high level and have a strong object-oriented flavor. They can be used to design systems, but lead to very large diagrams. Code can be produced.

The current Object Orientation programming languages (e.g., C++) are designed with concentration on the efficiency of the implementations and the continuation of coding practice from previous languages. The languages have the advantages of an algebraic (abstract data type) approach, but they are complex to use and coding is therefore error prone.

### Object Orientation and GIS

The advantage of Object Orientation methods for GIS have been seen early [8] and have been extensively discussed . Object Orientation can be used for the design and programming of GIS software; it is expected that this should improve the quality of the code, but is otherwise not relevant for the user. Object Orientation can also be used for the design of the programmers or the user interface and in this respect is highly relevant. New GIS designs -mostly from Europe- are based on object oriented concepts, which are available for the user or the application programmer. A number of research and R&D projects have been exploring different methods of building Object Orientation based interfaces to GIS, some using object oriented databases.

For the user of a GIS it is essentially irrelevant if the program is built with an Object Orientation programming language or not. One might expect that the software would be easier to maintain, but this is not guaranteed. For the designer of applications, it is useful if the application programmer interface has an Object Orientation. The object-oriented design of the GIS can show at the user interface. Several projects (Geo2, Geoworks) have produced consistent sets of specialized GIS operations.

### Current Efforts to Achieve these Goals

Activity today concentrates on the integration goal: how to build National Geographic Infrastructures and eventually also how to achieve a European GI. Standardization of data format has been mostly achieved. Lacking are process models.

### Integration of Data Using Interoperability

Standardization has progressed, a number of national and international data transfer standards are available and are used in parallel to the standard formats of the major vendors. Data transfer is less desirable, as it transfers a static snapshot of the data access using the rapidly expanding Internet to the updated data when they are needed is more attractive. The OpenGIS concept [5] and the related standardization allow access over the Internet independent of the software used to data, which is current.

Difficult today is the formalization of the description of the data (metadata), such that a potential user can find it and decide if the data another agency holds can be used to answer his questions. It is necessary to describe the phenomena the data describes, the quality of the data in general terms and the encoding. This is often summarized as meta-data (data describing the data) and closely linked to the semantics (meaning) of the data. It is widely recognized that describing meaning of data is closely related to the automated translation of natural language, which has been an elusive goal. To produce solutions, which can be used practically, solvable sub-problems must be addressed.

The current approaches to metadata, data quality and semantics are based on verbal (natural language) descriptions. Metadata typically takes the perspective of the data producer and describes- for lack of a better approach- the process used to collect the data. This is understandable only to a technically sophisticated potential user and does not lead to automatization required to wholesale integration of data from different sources as needed [30].

To achieve a formalization of metadata, the process of collecting the data and the process of using the data with respect to the real world must be modeled in a single framework. Then the correspondence between the phenomena in the world about which data are collected, and the phenomena in the world about which data are required to make a decision, can be linked and it becomes possible to decide on the fitness for use of the data automatically. This can be done for limited user communities, thus avoiding the need for a general solution of natural language understanding.

## Process Models

Dynamic models for processes are widely used in economics but also in environmental studies [29]. Most dynamic models are aggregate models, which model the change in parameters describing accumulated quantities in a system. Current modeling is also either concentrating on the physical aspects of the environment or on economic aspects.

Cellular automata allow models which show spatial distribution of quantities, but these models are typically very small. The University of Utrecht has recently demonstrated quite large and detailed models. Most recently, cellular automata models have been extended to include simplistic movable agents [9], which simulate behavior of human individuals in space.

Rapid advances in the raw computing power allow to increase the level of detail for such models. Missing are the software tools to routinely build and improve such models.

## Conclusions

The current GIS are systems to systematically collect, manage and present static data of the world. They present snapshots of what is. This limitation to a static view of the world limits their usefulness in many cases. It excludes GIS mostly from the limelight of the public debate. To help the politicians with the pressing questions of today, GIS must be extended to include dynamic data and the process models such that “what-if” questions can be answered.

In order to achieve this, the foundation of GIS must be extended from the current logic based framework of current (static) database management to include processes. Algebraic methods, which coincide with the object-oriented trend in software engineering, are perfect tools to model change and process.

Algebraic methods hold promise to allow the modeling of the semantics of data at least for limited user groups based on formal models of data collection and decision (data use) processes.

## Acknowledgements

Many of the concepts were developed during a long conversation with Alessandro Annoni in his hospitable house on the Lago Maggiore. Reading the book "Economy of Attention" by my colleague George Franck was instrumental in framing the question succinctly. Partial support for this research came from the JRC Ispra. We appreciate both the support and the stimulation. The opinions expressed are, of course, solely those of the author and not of the European Commission or the Joint Research Center. I appreciate the efforts of Mag. Roswitha Markwart and Isolde Schlaisich to improve this text.

## REFERENCES

1. Abel, D.J., et al. Support in Spatial Information Systems for Unstructured Problem-Solving, in Proceedings of the 5th International Symposium on Spatial Data Handling. 1992. Charleston: IGU Commission of GIS.

2. Albaredes, G. A New Approach: User Oriented GIS, in Proceedings of EGIS '92. 1992. Munich: EGIS Foundation.

3. Bird, R. and O. de Moore, Algebra of Programming. 1997, London: Prentice Hall.

4. Bittner, T. and A.U. Frank, On the Design of Formal Theories of Geographic Space. Geographical Systems, in press.

5. Buehler, K. and L. McKee, eds. The OpenGIS Guide - An Introduction to Interoperable Geoprocessing. 1996, The OGIS Project Technical Committee of the Open GIS Consortium: Wayland, Mass.

6. Buehler, K. and L. McKee, eds. OpenGIS Guide: An Introduction to Interoperable Geoprocessing, Part 1 of the Open Geodata Interoperability Specification (OGIS). 1996, The Open GIS Consortium, Inc. (OGC). Available from http://ogis.org/. 134.

7. Burrough, P. and I. Masser, European Geographic Information Infrastructures. I. Masser and F. Salgé, eds. Vol. 5. 1998, London: Taylor & Francis.

8. Egenhofer, M.J. and A.U. Frank. Object-Oriented Modeling in GIS: Inheritance and Propagation. in Proceedings of AUTO-CARTO 9, Ninth International Symposium on Computer-Assisted Cartography. 1989. Baltimore, MD: ASPRS/ACSM.

9. Epstein, J.M. and R. Axtell, Growing Artificial Societies. 1996, Washington, D.C.: Brookings Institution Press.

10. Eriksson, H.-P. and M. Penker, UML Toolkit. 1998, New York: John Wiley.

11. Frank, A.U., Socio-Economic Units: Their Life and Motion, in Life and Motion of Socio-Economic Units, J.-P. Cheylan, A.U. Frank, J. Raper, eds. to appear, Taylor and Francis: London.

12. Frank, A.U., J. Raper, and J.-P. Cheylan, eds. Life and Motion of Socio-Economic Units. ESF Series. to appear, Taylor & Francis: London.

13. Gallaire, H., J. Minker, and J.-M. Nicolas, Logic and Databases: a deductive approach. ACM, 1984. 16(2): pp. 153-184.

14. Goodchild, M., M. Egenhofer, and R. Fegeas, Conclusion Report of Interop'97, 1997, NCGIA.

15. Guenther, O., K.-P. Schulz, and J. Seggelke, eds. Umweltanwendungen geographischer Informationssystem. . 1992, Wichmann: Karlsruhe (Germany).

16. Konecny, M., ed. Geographic Information Systems: Information Infrastructures and Interoperability for the 21st Century Information Society (Proc. of GIS Brno'98 Conference, 28 June - 1 July, 1998). 1998, Masaryk University: Brno.

17. Kottman, C. Progress Toward Interoperability and a Geospatial Infrastructure at the Open GIS Consortium. in 4th EC GIS Workshop. 1998. Budapest: European Commission DG III.

18. McCarthy, J. and P.J. Hayes, Some Philosophical Problems from the Standpoint of Artificial Intelligence, in Machine Intelligence 4, B. Meltzer and D. Michie, eds. 1969, Edinburgh University Press: Edinburgh. pp. 463-502.

19. Meixner, H. and A.U. Frank, GI Policy - Study on Policy Issues Relating to Geographic Information in Europe. 1997, European Commission DG XIII: Brussels.

20. Morrison, J., The proposed standard for digital cartographic data. Amer. Cartographer, 1988. 15(1): pp. 9-140.

21. Navratil, G., An object-oriented approach to a model of a cadaster, M.Sc. thesis, Department of Geoinformation. 1998, Technical University of Vienna.

22. Neumann von, J. and O. Morgenstern, Spieltheorie und wirtschaftliches Verhalten. 1961, Würzburg: Physica-Verlag.

23. North, D.C., Institutions, Institutional Change and Economic Performance. The Political Economy of Institutions and Decisions, ed. J. Alt and D. North. 1997, Cambridge: Cambridge University Press.

24. Popper, K.R., Logik der Forschung. Die Einheit der Gesellschaftswissenschaften, ed. E. Boettcher. Vol. 4. 1984, Tübingen: J.C.B. Mohr.

25. Reiter, R., Towards a logical reconstruction of relational database theory, in On Conceptual Modelling, Perspectives from Artificial Intelligence, Databases, and Programming Languages, M. L. Brodie, M. Mylopolous, and L. Schmidt, eds. 1984, Springer Verlag: New York. pp. 191-233.

26. Ruth, M. and B.M. Hannon, Modeling Dynamic Economic Systems. Modeling Dynamic Systems. 1997, New York: Springer-Verlag.

27. Ryssevik, J., Long-Term Relationships Dealing with Boundary Changes when Analysing Aggregate Data, in Life and Motion of Socio-Economic Units, A.U. Frank, J. Raper, and J.-P. Cheylan, eds. to appear, Taylor & Francis: London.

28. Scholl, M. and A. Voisard, eds. Advances In Spatial Databases (5th Int. Symposium, SSD '97). Lecture Notes in Computer Science, Vol. 1262. 1997, Springer-Verlag: Berlin.

29. Thomas, R.W. and R.J. Huggett, Modelling in geography: a mathematical approach. 1980, New York: Harper & Row.

30. Timpf, S., M. Raubal, and W. Kuhn. Experiences with Metadata, in 7th Int. Symposium on Spatial Data Handling, SDH'96. 1996. Delft, The Netherlands (August 12-16, 1996): IGU.

31. Tomlin, C.D., Digital Cartographic Modeling Techniques in Environmental Planing. PhD thesis. 1983, Yale Graduate School, Division of Forestry and Environmental Studies.

32. Weibel, R., Map generalization in the context of digital systems. CaGIS, 1995. 22(4): pp. 259-263.

33. Worboys, M. A Model for Spatio-Temporal Information, in Proceedings of the 5th International Symposium on Spatial Data Handling. 1992. Charleston: IGU Commission of GIS.

34. Worboys, M.F., Object-Oriented Approaches to Geo-Referenced Information. International Journal of Geographical Information Systems, 1994. 8(4): pp. 385-399.

**Janet Franklin**
San Diego State University

Excerpted from -- Janet Franklin, Geographic Information Science and Ecological Assessment. In: Bourgeron, P., Jensen, M. and Lessard, G, *An integrated ecological assessment protocols guidebook*, Springer-Verlag, New York.

A growing number of GIS users -- ecologists, land use planners, and many others -- would probably profess (as one recently did to me) that a GIS is simply one of the data management tools they use -- in the same category as spreadsheet software. However, there are a number of institutional issues related to GIS and methodological issues related to spatial data analysis that distinguish it from other data storage support tools (spreadsheet, database software).

I think there is an uneven acceptance, or even awareness, or spatial perspectives in the other disciplines whose literature I read (ecology, forestry, wildlife biology, conservation biology, landscape ecology, remote sensing). On the one hand, sophisticated treatments by spatial analysts in ecology, etc., on the other hand this notion that spatial data are like any other data and GIS is a glorified database management system for holding those data. GISs have made spatial data widely available to physical scientists without necessarily making spatial analysis widely available.

**Mark Gahegan**
Department of Geographic Information Science, Curtin University of Technology, Perth, Western Australia

## The Application of Neural and Visual Techniques to the Analysis of Spatial Data

In recent times several developments in computer science have provided a wealth of opportunities for advancement of spatial analysis:-

1. Improvements in computing performance and the development of massively parallel architectures have enabled previously intractable analysis problems to be addressed via deterministic means.

2. Progress in pattern recognition, classification and function approximation tools, originating from the artificial intelligence community (such as decision trees, neural networks and genetic algorithms) now provide sophisticated capabilities for tackling a range of non-deterministic problems.

3. Advances in graphical display technology provide the basis for data exploration using visualisation or virtual reality techniques.

Taken together, these newer computing tools show considerable promise in that they are capable (in theory at least) of managing large, many-layered and heterogeneous datasets. This is just as well, since spatial analysis, whether concerned with the built or natural environments, has to deal with an increasing volume and diversity of data. Furthermore, with the advent of concern over global environmental issues, the scale and complexity of the tasks to be conducted is set to rise. However, many problems remain before this new technology is effectively harnessed. Many of these are methodological; sophisticated tools require sophisticated setup and operation.

**My main area of interest is in the development of suitable methods to make good use of these tools in a geographic setting, specifically for problems involving complex, high dimensionality datasets. Work to date has focussed on three different application areas: landcover classification (at the floristics level), spatial epidemiology and geological interpretation, and has led to the development of sophisticated visualisation tools and neural network-based classifiers**. Both of these areas are described briefly below.

### Visualisation for exploratory data analysis

Scientific Visualisation now provides the means to dynamically explore geographic datasets (Hearnshaw and Unwin, 1994) in a highly interactive, visual manner. As such it holds great potential as a tool for Exploratory Data Analysis (Haslett et al., 1991, Rheingans and Landreth, 1995), providing a collaborative working environment for knowledge discovery, data mining and hypothesis generation (all of which are poorly provided for in existing GIS). Figure 1 shows some example visual scenarios for exploring or hypothesising relationships between different environmental conditions and their associated vegetation.

It is important to note that the aim of exploratory visualisation is not to analyse the data *per se*, but rather to *present* the data to the user in a way that promotes the discovery of inherent structure and relationships (MacEachren & Ganter, 1990). In psychometric colloquialism this is known as inducing visual pop out (Csinger, 1992). Thus, a collaborative mode of interaction is developed between the user and the machine, where the visualisation environment produces a stimulus via the *visual encoding* of the data which is then *interpreted* by the user, enabling full advantage to be taken of the unsurpassed abilities of humans to perceive complex structural relationships.

Methods to achieve effective visual encoding strategies for spatial data have been investigated (Gahegan, 1996; 1998) including the use of expert knowledge to mediate and navigate through a combinatorially explosive range of possible solutions (Gahegan and O'Brien, 1997).

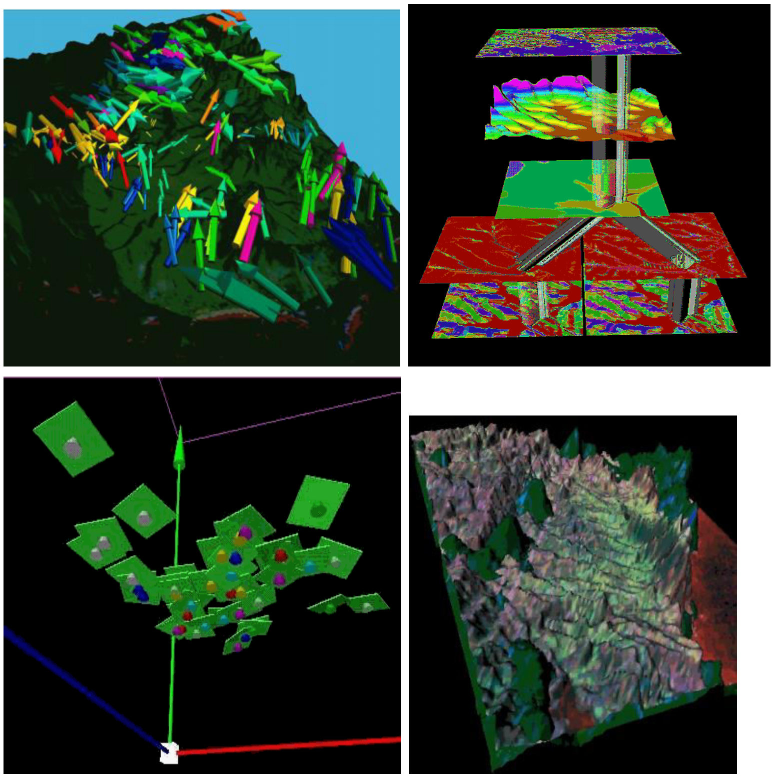


Figure 1. Four scenes depicting a range of visualisation techniques applied to the same data (an environmental dataset of a coastal region of New South Wales, Australia). Top left, mark composition using arrows draped on an elevation model. Top right, interactors describing relationships between different data layers. Bottom left, a scatterplot, enhanced with planar point icons to encode additional information. Bottom right, two different environmental surfaces are dynamically intersected. A comparison of these techniques can be found in Gahegan, 1998).

## Neural networks and decision trees for classification

The use of decision trees and (artificial) neural networks for data classification in geography and remote sensing has seen a steady rise in popularity. Kamata and Kawaguchi (1993) and Civco (1993) describe neural network classifiers whilst Lees and Ritman (1991), Eklund et al. (1994) and Freidl and Brodley (1997) describe classification approaches based around decision trees. Initially, the focus of attention was on comparing classifier performance with established methods (eg. Benediktsson, et al., 1990; Hepner et al., 1990; Paola and Schowengerdt, 1995; Fitzgerald and Lees, 1994). More recent efforts have concentrated on methodologies and customisation that improve performance or reliability; a sign that the technology has reached at least some level of acceptance. For example, Benediktsson, et al., (1993) and German et al. (1997) describe performance improvements and Kanellopoulos and Wilkinson (1997) and Gahegan et al. (1998) address methodological issues from the specific viewpoint of geographic datasets.

The kinds of classification problems that arise in geography or the wider earth sciences are often characterised by their complexity, both in terms of the classes and the datasets used. For example, classes may be difficult to define, may vary with location and over time, and their properties may overlap in attribute space. Datasets increasingly contain many descriptive variables (layers) and often contain a mix of statistical types; for example remotely sensed reflectance values (quantitative data) supplemented with nominal data such as soil type or geology and ordinal data such as slope or aspect. Data saturation seems set to increase with the adoption of sensing devices of greater sophistication, resulting in a higher spatial resolution and many more channels. The complexity of the tasks to which these data are applied is also increasing; for example the classification of deep geological structure from 300 channel airborne electro-magnetics data, or socio-demographic indices from combinations of many indicator variables.

Addressing geographic classification problems successfully with tools based around inductive learning and search requires detailed attention to methodology and often also a good deal of further development and enhancement. To this end a black-box neural classifier has been specifically engineered for application to geographic problems (German and Gahegan, 1996). It is based on a feedforward, multi-layer perceptron, with various enhancements made (German et al., 1997; Gahegan et al., 1998) Importantly, it is designed to be self-configuring, requiring only the same setup as a standard Maximum Likelihood Classifier. Results so far show increased classification accuracy over established techniques, that is maintained as the number of data layers (attributes) increases to twenty or more.

## References

Benediktsson, J. A., Swain, P. H. and Ersoy, O. K. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. IEEE transactions on Geoscience and Remote Sensing, Vol. 28, No. 4, pp. 540-551.

Benediktsson, J. A., Swain, P. H. and Ersoy, O. K. (1993). Conjugate gradient neural networks in classification of multisource and very high dimensional remote sensing data. International Journal of Remote Sensing, Vol. 14, No. 15, pp. 2883-2903.

Civco, D. L. (1993). Artificial neural networks for landcover classification and mapping. International Journal of Geographical Information Systems, Vol. 7, No. 2, pp. 173-186.

Csinger, A. (1992). The psychology of visualisation. Technical Report Series, Department of Computer Science, University of British Columbia, Canada.

Eklund, P. W., Kirkby, S. D. and Salim, A. (1994). A framework for incremental knowledge update from additional data coverages. Proc. 7th Australasian Remote Sensing Conference, Melbourne, Australia, Remote Sensing and Photogrammetry Association of Australia, pp. 367-374.

Fitzgerald, R. W. and Lees, B. G. (1994). Assessing the classification accuracy of multisource remote sensing data. Remote Sensing of the Environment, Vol. 47, pp. 362-368.

Freidl, M. A. and Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. International Journal of Remote Sensing, Vol. 61, No, 4, pp. 399-409.

Gahegan, M. N. (1996a). Visualisation strategies for exploratory spatial analysis. Proc: Third International Conference on GIS and Environmental Modelling, NCGIA, Santa Barbara, USA.

Gahegan, M. (1998). Scatterplots and scenes: Visualisation techniques for exploratory spatial analysis, Computers, Environment and Urban Systems, Vol. 22, No. 1. (in press).

Gahegan, M., German, G. and West, G. (1998). Some solutions to neural network configuration problems for the classification of complex geographic datasets. To appear in Geographical Systems.

Gahegan, M. N. and O’Brien, D. L. (1997). A strategy and architecture for the visualisation of complex geographical datasets. International Journal of Pattern Recognition and Artificial Intelligence, Vol. 11, No. 2, pp. 239-261.

German, G. and Gahegan, M. (1996). Neural network architectures for the classification of temporal image sequences. Computers and Geosciences, Vol. 22, No. 9, pp. 969-979.

German, G., Gahegan, M. and West, G. (1997). Predictive assessment of neural network classifiers for applications in GIS. Second International Conference on GeoComputation, Otago, New Zealand, pp. 41-50.

Haslett, J., Bradley, R., Craig, P., Unwin, A. and Wills, G. (1991). Dynamic graphics for exploring spatial data with application to locating global and local anomalies. The American Statistician, Vol. 45, No. 3, pp. 234-242.

Hearnshaw, H. M and Unwin, D. (Eds.) (1994). Visualization in Geographical Information Systems, John Wiley & Sons, Chichester, England.

Hepner, G. F., Logan, T., Ritter, N. and Bryant, N. (1990) Artificial neural network classification using a minimal training set: comparison to conventional supervised classification. Photogrammetric Engineering and Remote Sensing, Vol. 56, No. 5, pp. 469-473.

Kamata, S. and Kawaguchi, E. (1993). A neural network classifier for multi-temporal Landsat images using spatial and spectral information. Proc. IEEE 1993 International Joint Conference on Neural Networks, Vol. 3, pp. 2199-2202.

Kanellopoulos, I. and Wilkinson, G. (1997). Strategies and best practice for neural network image classification. International Journal of Remote Sensing, Vol. 61, No, 4, pp. ???

Keller, P. R. and Keller, M. M. (1993). Visual Cues: Practical Data Visualization, IEEE Press, Los Alimatos, CA, USA.

Lees, B. G. and Ritman, K. (1991). Decision tree and rule induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. Environmental Management, Vol. 15, pp. 823-831.

MacEachren, A. M. and Ganter, J. H. (1990). A pattern identification approach to cartographic visualisation. Cartographica, Vol. 27, No. 2, pp. 64-81.

Paola, J. D. and Schowengerdt, R. A. (1995). A detailed comparison of backpropagation neural networks and maximum likelihood classifiers for urban landuse classification. IEEE transactions on Geoscience and Remote Sensing, Vol. 33, No. 4, pp. 981-996.

Rheingans, P. and Landreth, C. (1995). Perceptual principles for effective visualisations. In: Perceptual Issues in Visualisation, (Eds. Grinstein, G. and Levkowitz, H.), Springer-Verlag, Berlin, pp. 59-69.

**Jacqueline Geoghegan**
Department of Economics, Clark University, Worcester, MA

Research on the human dimensions of global change reveals the urgent need to insert nature-society and spatial relationships more fully into social science problem solving and theory development. One potentially important approach is to explore the usefulness of linking remotely sensed data and geographical information systems (GIS) to the core social science themes embedded in the ‘human dimensions’. Yet, such linkages have been explored slowly, even reluctantly, by social science. The reasons for this low level of receptivity are many. RSD are thought to be peripheral to the core themes of the social sciences and potentially invade confidentiality of the source being studied (NRC, 1998). Also, owing to data barriers, the social sciences have emphasized aspatial conceptualizations of social structures, processes, and decision making that de-emphasize the very strengths that RSD and GIS bring to problem solving. When faced with a fundamentally spatial question, social science approaches commonly abstract away from its spatial nature, or when space is explicitly incorporated into social science models, it often enters simply as a constraint on the system or process under investigation.

As the emergence of high spatial resolution (e.g., TM or SPOT) and high temporal resolution (e.g., NOAA AVHRR/METEOSAT) satellite imagery afford opportunities to explore nature-society relationships at spatial scales consistent with social science theory and concepts, as well as the creation of other social science spatial databases, and as advances in GIS provide unprecedented abilities to analyze these data more social scientists are beginning to use these data and technologies. But there still remains a fundamental question for the social sciences: will RSD and GIS affect problem solving and theory development in the social sciences as profoundly as they have the natural and, to a lesser degree, applied sciences?

The challenges to the social sciences are theoretical, empirical, and methodological. For example, now that spatial data are increasing, a series of issues about the methods to best make use of them must be addressed. The subfield of spatial econometrics, for example, is rapidly expanding to meet the needs of modelers, but much more research is needed on spatial estimation tools to test the hypotheses derived from theory. I am a Principle Investigator or Co-Principle Investigator in two large projects, that are involved in an interdisciplinary framework to advance this frontier focused on land-use/land-cover change.

The first is the Southern Yucatán Peninsular Region (SYPR [NASA-LCLUC]) project in cooperation with Harvard Forest and El Colegio de Frontera Sur (Mexico). This interdisciplinary project aims: to understand, through individual household survey work, the behavioral and structural dynamics that influence land managers' decisions to deforest and intensify land use; model these dynamics and link their outcomes directly to TM imagery through GIS; model from the imagery itself; and, determine the robustness of modeling to and from the RSD. Several critical social science themes are addressed: how can decisions based on market and subsistence be integrated in one model; how robust are decision based models in the face of volatile, exogenous forces; and what is the value added of using GPS and GIS in survey research to investigate how the individual chooses land-use practices and how these explicitly vary over space and time.

The second interdisciplinary project, the Patuxent Watershed of Maryland (EPA/NSF) collaborative with the University of Maryland, focuses on the links and feedbacks between human decisions on land development and ecological consequences. Econometric model predicts the probably of land-use change in the watershed as a function of both economic and ecological spatial variables. These economic models, when linked with any number of ecological models, allows the effects of both direct land-use change through human actions and indirect effects through ecological change to be evaluated. This project is unique in that it is the most spatially explicit and dissaggregated model of individual human behavior that currently merges RSD and GIS in economic modeling.

For a fuller description of these two projects, see Geoghegan et al., (1998). Central to the framework of both of these projects is the attempt to insert human spatial and human-environment processes into our analysis by way of RSD and GIS, not only to provide insights about spatial outcomes but to inform and evaluate the basic theoretical concepts underpinning the substantive questions.

**References**

Geoghegan, Jacqueline, Lowell Pritchard, Jr., Yelena Ogneva-Himmelberger, Rinku Roy Chowdhury, Steven Sanderson, and B. L. Turner II, 1998. 'Socializing the Pixel' and 'Pixelizing the Social' in Land-Use and Land-Cover Change. In People and Pixels: Linking Remote Sensing and Social Science. Committee on the Human Dimensions of Global Environmental Change, National Research Council. 1998.

National Research Council, 1998. People and Pixels: Linking Remote Sensing and Social Science. National Academy Press, Washington DC.

**Arthur Getis**
Department of Geography, San Diego State University

For the further development of local statistics, it is necessary to better understand their meaning when the underlying global autocorrelation is not zero and when a series of dependent tests are employed. The problem is to find a test for local spatial autocorrelation in the presence of global spatial autocorrelation. The practical goal is to find a means for identifying statistically significant ‘hot spots’ or clusters in data presented either in raster or vector form on two dimensional surfaces.

A fundamental concern of pattern researchers is to find peculiarities in spatial data that lead to the identification of hot spots or clusters that signify that something out of the ordinary has occurred in one or more regions within the area covered by the data. In fields such as regional science, economic geography, and epidemiology questions are often raised about the number of events in particular subregions of the study area. These include concerns such as the location of migration destinations, the location of specialized economic activity, and the location of infectious diseases. In recent years, a number of papers have been written where the identification of hot spots is the principal concern.

A natural direction for this research has been to develop statistics that can pinpoint the exact location of places that exhibit these special characteristics. Previous to the development of local statistics, one depended on indicators such as the mapped residuals from regression to identify spatial outliers. Well known global statistics such as Moran’s I, Geary’s c, and Matheron’s variogram are not designed to look beyond the general autocorrelation characteristics of a pattern, although they can be made sensitive to directional influences. Cressie (1991), for his pocket plot, while identifying outliers, does not provide a test for statistical significance. The variogram does dissect patterns into their component correlations by distance increments as do the correlograms based on Moran’s and Geary’s statistics, but these do not depend on a single spatial focus as do local statistics.

The local statistics of Ord and Getis (GA: 1992, 1995) and the LISA statistics of Anselin (GA: 1994) are designed to test individual sites for membership into clusters. Both Ord and Getis and Anselin recognize, however, that in the face of global autocorrelation, finding individual centers of clustering becomes a problem. Ord and Getis provide a discussion of the issue with a proof that local statistics must be interpreted differently at different levels of global autocorrelation.

Ord and Getis also addressed the issue of finding appropriate statistical test cutoff values of their local statistics when multiple simultaneous dependent tests are employed. This is an issue of some importance, especially when the technology is moving us in the direction of larger data sets.

**Daniel A. Griffith**
Department of Geography, Syracuse University, NY

## HOW UNSUCCESSFUL HAS GIS BEEN IN HELPING DISSEMINATE SPATIAL STATISTICAL TECHNOLOGY?

Implementation of spatial statistical techniques has been problematic since their initial appearance in the academy. Like GIS databases, spatial statistical analysis is hampered by a need to retain spatial structure (the map to which georeferenced data are linked as well as selected analytical features of this structure an ability absent in most standard commercial software packages implementing traditional statistical techniques. Exceptions include, to a limited degree, S+ (the spatial statistical module) and more recently SAS (PROC GIS). Attempts to circumvent this restriction mostly have resulted in dedicated software, such as SAGE (for spatial autoregression, built upon ArcInfo), SpaceStat (with interfaces to GIS packages), and GSLIB (for geostatistics), although some efforts have tricked standard packages into executing spatial statistical procedures (e.g., Griffith's SAS and MINITAB code developments). The importance of remedying this situation is attested to by an emphasis on the need for promoting spatial statistical skills in such expert statements as the UCGIS education white paper. The question addressed in this paper asks whether or not GIS has helped, is helping, and/or will be helping disseminate spatial statistical technology in order to promote the acquiring of spatial statistical skills by members of the GIS community. Not surprisingly the answer to this question comprises a mixture of yeses and nos.

Underlying the recognition of a need for user-friendly and easy access to spatial statistical technology is an appreciation of what is special about georeferenced data. In other words, what is spatial autocorrelation and why should a spatial scientist not overlook this property of georeferenced data? GIS software holds a privileged position for aiding students in responding to this question. Certainly one simple way of addressing this issue is to provide a tutorial in a GIS that allows a user to discover the answer to this question. Such a tutorial could be modeled after SASIM, EXPLORHO, or USA, for example. Another possibility is to design a tutorial that focuses on the interpretation of spatial autocorrelation as redundant information: the objective would be for a student to decide whether or not a superfund site should undergo remediation, and if so, what sections should be remediated. One of the fascinating features of this illustrative context concerns determination of the effective sample size, which virtually always will be much less than n (the number of sample points); in the extreme, if a student selects all points for a sample from essentially the same location, the effective sample size will be only slightly greater than 1! Such a dramatic outcome is highly effective in revealing what is special about spatial data. To date, GIS software at best only allows a user to compute standard spatial autocorrelation indices (Moran Coefficient, Geary Ratio), which sheds little light on this practical issue.

A second part of the question being addressed here concerns spatial structure and its selected analytical properties necessary for executing a spatial statistical analysis. GIS database structures already furnish a means for retaining the necessary geographic structure when a spatial scientist is interested in conducting spatial analysis. This explicit recording of geographic structure can be exploited in order to implement spatial statistical techniques (see, for instance, Zhang and Griffith, 1998a,b). But the selected analytical features needed are eigenfunctions, a concept foreign to or perceived as being highly intimidating by most students of the spatial sciences. A GIS offers a means of redressing this situation. First, spatial statistical techniques can be implemented in a way that requires knowledge of only the extreme eigenvalues characterizing a given geographic structure. Such an implementation alleviates considerable computational and computer memory burdens, and allows massively large georeferenced datasets to be analyzed. Theoretical spatial statistical results already exist for implementing this approach; what remains is for GIS vendors to incorporate these results. In addition, preliminary research suggests that even the extreme eigenvalues may remain unknown; this research needs to be finalized.

Second, the eigenvectors characterizing a geographic structure reveal a kaleidoscope of possible distinct levels and patterns of spatial autocorrelation. A tutorial needs to be developed for inclusion in a GIS that exploits this feature. Pedagogically speaking, it holds considerable promise for sharpening the map-reading skill of inspecting a map pattern and intuitively being able to ascertain the approximate nature and degree of spatial autocorrelation present in the visualized geographic distribution. Experimentally speaking, it holds considerable promise for resampling experiments in which specific levels of nonzero spatial autocorrelation are to be explored.

While a number of GIS packages have embraced spatial autocorrelation indices, to date somewhat more attention seems to have been devoted to geostatistical implementation. In general, though, everyone seems to have begged the question asking what the relationship is between spatial autoregression and geostatistics. Spatial scientists and students alike continue to puzzle over this question. A GIS environment is the perfect one to foster a far more comprehensive and deeper understanding of this relationship, since both approaches to georeferenced data analysis focus on the property of spatial autocorrelation. Of course, in order to do so spatial autoregression techniques first need to be implemented in GIS packages.

Therefore one answer to the question addressed by this paper may be summarized as follows:

GIS has been, and continues to be, somewhat successful in raising the awareness of spatial scientists and students about selected features of geostatistics and popular indices of spatial autocorrelation, but overall continues to be quite unsuccessful at making spatial statistical technology widely available to this same group. It certainly possesses the potential to do so, however!

Accordingly, one future research endeavor should be to: introduce tutorials into GIS packages that teach about spatial autocorrelation, and implement spatial autoregression procedures in GIS packages, completing these two tasks in such a way that a user can gain a better understanding about the relationship between geostatistics and spatial autoregression. These achievements should be guided by and would reflect upon the decade-old debate concerning what spatial analysis routines should be incorporated into a GIS. Moreover, while the GIS environment appears to be satisfactory, implementation of spatial statistical procedures remains neglected by GIS vendors, with one result being a failure for spatial scientists to have access to state-of-the-art research methodology

**Stephen C. Guptill**
U. S. Geological Survey, Reston, Virginia

## Spatial Analysis of Vector Borne Diseases

For many hundreds of years people have been intuitively aware of the relationships between human health and the environment. Today, geographic information systems, remote sensing satellites and other technologies are providing scientists with the tools and the data to make clear the geographic relationships between environmental habitats of disease vectors and agents and the occurrence of disease. By knowing the geographic conditions necessary for the maintenance of specific pathogens in nature, one can use the landscape to identify the spatial and temporal distribution of disease risk. Key environmental elements (including land cover, elevation, temperature, and rainfall) influence the presence of pathogens, vectors, zoonotic reservoirs of infection, and their interactions with humans. A program of joint research is being developed with scientists at the Centers for Disease Control and Prevention to examine the environmental influences on Lyme disease, plague, and viral encephalitis. Specifically we are working to:

- Define geographic distributions of disease cases and relationships to environmental factors.
- Develop and test a model predicting disease activity and transmission rates.
- Characterize the human population at risk.
- Devise ecology-based prevention and control measures.

However, the analytical capabilities available in ‘off the shelf’ GIS packages are inadequate to address these tasks. This research requires the use of a 4D GIS model (with time sensitive variables such as temperature, rainfall, and humidity), the use of probabilistic models, as well as spatial statistical tools. While some capabilities can be assembled from existing software (e.g. S-Plus), others will need to be custom built. The need to construct an adequate spatial analysis environment will delay us in conducting our epidemiological studies.

**Robert P. Haining**
Department of Geography, University of Sheffield, Sheffield, UK

Following many years researching in the area of spatial statistical analysis I collaborated with Stephen Wise (Univ of Sheffield) in the development of SAGE - a software system that provided spatial statistical data analysis capability within a GIS (Arc/Info) environment. The system included exploratory and confirmatory spatial analysis techniques in a linked windows environment. The system uses table, graphs and map windows. The aim of the project (funded over three years by the ESRC) was to develop a system that provided proper spatial statistical analysis capability within a GIS, exploring in the process what developments were possible. The project finished in 1997, papers are now appearing and additional papers are in preparation demonstrating the application of SAGE to health and crime data analysis.

The research raised interesting issues of of how visualisation can be used to assist exploratory and other forms of spatial data analysis. Haining and Wise obtained a small grant at the end of 1997 from ESRC/JISC to carry out a comparative analysis of software systems and a paper is due out in E&P A reporting that assessment. A further paper was presented at this years ERSA describing a conceptual framework for the development of visual tools for exploratory spatial data analysis (ESDA) and using this to critique SAGE.

My contribution to the workshop would hinge on the development of SAGE as an interactive software system for ESDA but focussing on the development of visualisation tools. We see the development of such tools as critical in the sense of allowing wider participation in the process of extracting information from spatial data. Such a process need not be the exclusive domain of the ‘spatial statistical specialist’ nor the domain of the ‘automated pattern spotter’. Using intuitive but informative tools it becomes possible to engage a wide range of interested parties in seeking to tease out important information from large data sets. We already have some limited experience of this with Sheffield Health in constructing deprivation regions for Sheffield.

**Geoffrey M. Jacquez**
BioMedWare, Ann Arbor, Michigan

## Placing the Horse in Front of the Cart: Spatial Analysis and GIS

### Background and Terminology:

I assume `spatial analysis' means spatial statistical analysis, rather than GIS operations such as overlay, buffering and etc. that are called `spatial analysis' by some practitioners. I assume spatial pattern in a variable/map is the consequence of past physical, biological, sociological, and geological processes. Finally, I assume our goal as scientists is to increase our understanding of these forces, and we analyse spatial pattern in order to infer their past action. The contribution of spatial statistical analysis is to (1) detect spatial pattern, and (2) determine whether a pattern is significant and thus merits formulation of explanatory hypotheses.

### Problem Statement:

GIS technology leads the science, and arcane software design considerations underlying legacy GIS often dictate the scientific issues we address. Spatial statistical analysis in GIS is an afterthought, and it therefore seems reasonable to suppose that current GIS are not particularly good platforms for incorporating tools for spatial statistical analysis. What if we went through the GIS design process using spatial statistical analysis of geodata as the objective? Such a Spatial Statistical GIS (SS-GIS) might incorporate the following characteristics:

- The ability to construct 'designer' spatial statistics appropriate for the available data and the particular problem at hand.
- Spatial queries appropriate for quantifying proximity metrics (e.g. spatial weights) required by spatial statistics.
- Mechanisms for modeling location uncertainty and for quantifying the spatial sampling space.
- Spatial Monte Carlo methods supporting restricted null hypotheses other than Complete Spatial Randomness (CSR).

### Statement of Interest:

My colleagues and I have developed a prototype SS-GIS (called Gamma) that addresses these concerns. It uses a flexible mathematical form (called the gamma product) for representing spatial statistical tests, and spatial Monte Carlo techniques that provide a common mechanism for assessing statistical significance. Monte Carlo randomization may be restricted to account for spatial dependency (e.g. spatial autocorrelation under the null model). An equation editor allows one to create custom multivariate data metrics (e.g. calculated from several attributes). Three location models (point, polygon, population) may be used to specify the spatial sampling space, and to model location uncertainty. I am interested in attending the meeting to share these results with other researchers; to learn of recent advances; and to put in my two (or three) cents in the `stock-taking' that is the purpose of the workshop. This workshop has the promise of identifying design requirements of an `SS-GIS’ focused on meeting researchers needs. Will we finally put the horse before the cart and have the science lead the software?

**Konstantin Krivoruchko**
Environmental Systems Research Institute, Redlands, California

# GIS and Geostatistics: Spatial Analysis of Chernobyl's Consequences in Belarus

## Abstract

A large amount of data concerning the ecological state of Belarus after the Chernobyl accident was analyzed using geostatistics. This data included soil contamination by long-lived radioisotopes $^{137}$Cs, $^{90}$Sr, $^{238-240}$Pu, and $^{241}$Am; dose load estimates on the population during the first days after the Chernobyl accident; $^{137}$Cs food contamination and estimation of the internal dose in 1993; and thyroid cancer morbidity among adults and children based on data collected from 1986 to 1995.

Currently, in order to integrate powerful methods of spatial data interpolation into a geographic information system (GIS), users are required to use statistical packages to process and then save the results of interpolation into a GIS-supported format. To analyze complicated environmental problems and to present the results of an analysis on a map requires a group of specialists who are familiar with both GIS techniques and complicated statistical software. In this article we discuss an approach of combining advanced spatial statistics and GIS. This combination was first presented in the GIS MapStudio (Krivoruchko et al., 1997; Maignan, Krivoruchko, 1997) software package, which integrated both modern visualization techniques and geostatistical methods of spatial data analysis. These methods are currently being considered for further development and implementation into ESRIs GIS software line.

This paper highlights the advantages of using geostatistical methods for processing environmental data. The following geostatistical approaches are used for mapping: simple, ordinary, indicator, probability, and disjunctive krigings, as well as ordinary and indicator kriging modifications for binomial data. The applicability of the different data processing methods is discussed. The data and the visualization techniques presented in this paper can help reveal powerful patterns for decision making and policy planning.

## Introduction

The Chernobyl accident affected Belarus more than any other country. Ironically, Belarus does not have a single nuclear power plant; however, a ring of nuclear power stations surrounds it including the Ignaline station in Lithuania, the Smolensk station in Russia, and the Chernobyl station in Ukraine. The wind-roses on April 26-30, 1986, was such that about 70 percent of the radioactive dust from Chernobyl fell on Belarus. The analysis and visualization of environmental and epidemiological data, therefore, are critical for predicting public health issues. Many interpretations of the effects from the

Chernobyl accident appeared in post-Chernobyl years. The effects of radiation 12 years following the accident, however, are yet to be fully explored and are therefore poorly understood. The situation has been exacerbated by a lack of experience using modern statistical interpolation methods together with a GIS. Such techniques can reveal interrelationship between spatial distributions of radioactive contamination and disease incidence rates.

The Chernobyl accident released about $1.85 \times 10^{18}$ Becquerel of radioactive material. Each Becquerel represents one radioactive decay event per second. In 1958, Andrey Sakharov described "nonthreshold effects," by which every radioactive particle released had a statistical probability of doing damage either to the DNA of a cell or to the immune system through low-level internal radiation caused by ingesting such particles. He also predicted that radiation would accelerate the mutation of microorganisms, leading to the inference that persons with damaged immune systems would in time succumb more easily to these new strains (Sakharov, 1958). The epidemiological studies carried out in Belarus have revealed an increase in the incidence of thyroid and lung cancers, cardiovascular pathologies, and pregnancy anemias and an increase in mortality due to birth defects occurring in the contaminated regions (Lutsko et al., 1996).

During the first days after the accident, residents of Belarus absorbed the majority of their radiation dose through their thyroid glands via inhalation of contaminated air and, more importantly, consumption of contaminated foodstuffs, mainly cow's milk. Levels of contamination by short-lived radionuclides, in particular $^{131}$I, were so high that the corresponding exposure of millions of people was qualified as "iodine shock". This will result in long-term health problems for the population. The epidemic of childhood thyroid cancer is the first indisputable long term health aftereffect of the accident.

At present (and over the next decades), the main hazard comes from $^{137}$Cs ($^{90}$Sr, $^{238-240}$Pu, and $^{241}$Am also play a significant role at short distances from the reactor). Today, internal exposure from intake of $^{137}$Cs contaminated food contributes to more than half of the whole radiation dose received by the population. Income of the inhabitants of villages in Belarus does not afford them access to "clean" (non contaminated) food. They consume vegetables, potatoes, and milk produced on their own contaminated personal properties as well as mushrooms and berries from nearby forests.

The main goal of investigating the Chernobyl accident is to estimate the risk of health hazard to the population, which was irradiated and continues to live in the contaminated territory, based upon information about the total exposure of the population. In the current article we attempt to:

Reveal the interrelationship between the spatial distribution of cancers based upon radionuclides soil contamination and population dose loads.

Use different methods for processing spatial data, describing briefly its accuracy and peculiarities.

**Methods of Data Analysis**

Geostatistics are statistical methods used to describe spatial relationships among sample data and to apply this analysis to the prediction of spatial and temporal phenomena. They are used to explain spatial patterns and to interpolate values at unsampled locations. Geostatistics have traditionally been used in the sphere of geosciences: meteorology, mining, soil science, forestry, fisheries, remote sensing, and cartography. Geostatistical techniques were originally developed by Soviet scientists for meteorological data predictions. The first book with complete explanations about simple and ordinary kriging and cokriging techniques was published in Leningrad in 1963 (Gandin, 1963). According to this book, the original name of the technique is*objective analysis*. Geostatistical techniques were later successfully applied to mining and other disciplines.

In addition to the description of spatial patterns and interpolation of data, important components of geostatistical analysis include the integration of secondary data in prediction algorithms through the use of cokriging and error analysis. The cokriging technique allow one to compute an optimum estimation while considering the relationships between several associated variables. Using error analysis, validation, and cross-validation, it is possible to assess the performance of the interpolation as well as to reveal errors in the source data.

Sampling and mapping in the earth sciences are complicated by complex spatial and temporal variations. The structure and intensity of patterns being sampled often cannot be determined or predicted reliably with deterministic models because of uncertainties in the data and the phenomena under investigation. The best we can do using interpolation and estimation methods is to be as objective as possible and to consider the interrelationship of the data under investigation.

Deterministic approaches to interpolation (e.g. trend surface, inverse distance weighting, triangulation, and splining) are based upon *a priori* mathematical models of spatial variation and can lead to smooth but inaccurate maps. This is because error is an inherent part of the sampling process. In practice, error can not be eliminated but only minimized. Therefore, in most cases one cannot produce a representative map of estimated values in unsampled locations with these techniques.

Geostatistical processes are comprised of the three major components of regionalized variable analysis: variographical analysis, prediction making, and error analysis.

During structural analysis, spatial (auto)correlation can be analyzed using covariance and semivariogram. After structural analysis, predictions at unsampled locations are made using kriging (i.e. transposition of multiple linear regression into a spatial context).

The value of the error variation shows the uncertainty of the data at the considered point. Kriging variance is higher in poorly informed zones. Simple and ordinary kriging variance has a limited usefulness in the case of nonGaussian distribution. In this case the map of error of estimation associates with the data density and kriging variance equals to average variance of the selected neighborhood. If normal score transform had used and the bivariate normality was adopted, we can correctly calculate the confidence interval of estimation based on the features of the normal distribution. Thus, map of error estimation will provide an accurate measure of the local estimation precision.

This fact makes it possible to obtain maps of the probability of exceeding some predicted level by the variable under investigation (i.e., kriging provides the user with a tool that can help in the decision making process) as well as to draw the isolines with fixed significance.

For producing risk-qualified maps, the nonparametric geostatistical algorithms such as indicator kriging, indicator cokriging, and probability kriging are primarily recommended. They were developed to process data of highly variant phenomena without having to trim off important high-valued data.

Validation allows one to randomly select parts of the database, perform kriging, and then compare results of estimation with true values. One more procedure of error analysis, in addition to validation and mapping errors of estimation or errors of indicators, is cross-validation. This method consists of analyzing and displaying estimated values calculated under the assumption that particular sample data is missing.

Cross-validation allows one to find regions where data is over/underestimated and to compare the validity of different geostatistical approaches for data interpolation. By comparing sample values and their estimations, one can reveal data points having the greatest standardized errors. Such values demand additional attention because they may be data outliers. The result of validation and cross-validation techniques can be analyzed further by using graphics tools and spatial analysis functions.

Some additional tools are available for the geostatistical data analysis. In this article we used declustering techniques, which allows one to improve estimation if measurements are sampled preferentially (see for example Deutsch, Journel, 1998).

In the article we used the following geostatistical approaches: *simple, ordinary, indicator, probability and disjunctive krigings*as well as *ordinary and indicator kriging modifications for binomial data*.

*Simple* (in this method the estimation of the mean can be established *a priori* based upon a different data set from the data used for the present estimation; for example, declustering mean) and *ordinary* (with unknown mean) krigings (Gandin, 1963) can be applied successfully to data having, or close to, a multivariate Gaussian distribution. The assumption of the multivariate Gaussian distribution is rarely realized in practice. To check for the multivariate normal distribution of the data, one should first transform the initial data to a univariate normal distribution (i.e. performing linearization of the original data). Such transformation performed prior to the variogram/covariance analysis allows one to reduce the variability in the original data, making the variogram modeling more reliable and stable. The normal score transform function can be defined for any continuous cumulative distribution function. Transformation of the distribution of the initial data set into univariate Gaussian distribution does not guarantee that all of the initial data will be transformed into an exact multivariate Gaussian distribution. It is, however, relatively easy to check for bivariate normality of the transformation (Deutsch, Journel, 1998). If bivariate distribution is fulfilled, transformation could be adopted. Simple and ordinary kriging provide an optimal estimation in the class of linear models. Suggestion about normality of the transformed data allows one to calculate the confidence interval for the estimation and to present results of the estimation as a map with the desirable level of significance and as a map of the probability that selected critical level is exceeded or not, in addition to the map of estimation.

*Indicator* kriging (Journel, 1988) is an example of the techniques in which the uncertainty model depends only on the information available. In indicator kriging a 0-1 transformation is introduced, which will be the new variable for which we will compute variograms and carry out kriging methods. Among the shortcomings of *indicator* kriging is the loss of information after coding data through indicator functions. For example, if the data values are in the range of 1 to 100 and the indicator value (cutoff) selected is 20, then the data points with values 21 and 99 will be interpreted as being equivalent. It is possible, therefore, in such situations to use a soft indicator function. This indicator function can be prepared by the user based upon knowledge of the data sets and processing by ordinary kriging. One of the possible solutions is to remove data near the threshold from consideration.

*Probability* (Journel, 1988) kriging is considered an improvement over the indicator kriging method in the sense that the data is used more completely for the task of estimating conditional probabilities that given thresholds are exceeded. This nonparametric approach uses the original data set in addition to indicators.

*Disjunctive* kriging (Rivoirard, 1994) is also based on a specific nonlinear transformation of the original data. The first step in this type of estimator is the normal score transformation of the data. The next step is to express the normalized function as an expansion of hermitian polynomials. It should be noted that a sufficient number of hermitian polynomials depends on searching neighborhoods and to receive reliable results of estimation we should use a different number of hermitian polynomials for different locations. It is also assumed that the condition of bivariate standard normal distribution is fulfilled. It should be noted that bivariate normal distribution is weaker than the multi-Gaussian condition. As mentioned above, it is possible to check the bivariate normality of the transformation. If bivariate normality exists, disjunctive kriging can produce better results than other geostatistical methods. If it is not so, it is desirable to use another type of kriging. Validation and cross-validation techniques can be used to confirm this.

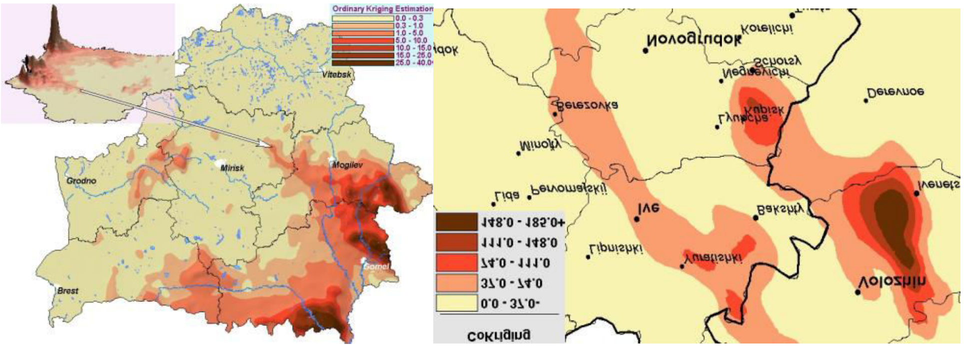
Environmental applications often include a collection of data, that can be processed by *cokriging*. Cokriging is the logical extension of kriging to situations where two or more variables are spatially interdependent and where the one whose values are to be estimated is not sampled as intensively as the others with which it is correlated. In other words, cokriging is used when one variable is of principal interest and the other variables are used to enhance the estimation of the first (i.e. the supplementary variable is used to compensate for a lack of data for the primary variable). Analysis of simple and cross-semivariograms, which are created during this process, allows one to check the spatial correlation between different variables. In general, using cokriging techniques will always result in more precise estimations. The influence of the secondary data set over the primary data set is higher if the primary data set includes the greatest error and if correlation between the data sets occurs. If the correlation between the data sets is negligible, we will receive the same result as kriging one variable. The difference of data estimation based on kriging and cokriging approaches could also help to find outliers.

### Long-Lived Radionuclides Soil Contamination

Since the accident at Chernobyl, public safety in the territories with long-term radioactive contamination has assumed a fundamental importance. Investigation of soil radioactivity is a necessary step in the detailed description of the sources of radioactivity exposure because contaminated soil is the main supplier of radioisotopes into other components of the biosphere. This is the only possible way at the present time of determining the hazard from $\alpha$ -radiators (mainly $^{238-240}$Pu and $^{241}$Am) to human health since the direct dose measurements in this case are infeasible.

According to *International Basic Standards for Protection against Ionizing Radiation and for the Safety of Radiation Sources* (International, 1996), radiation dose limits in unrestricted (uncontrolled) areas shall be such that an individual will not receive a dose to the whole body in excess of *1 millisievert (mSv) per year*. The dose is the cumulative exposure to radiation actually received by an individual. In practice, however, classifications of the contaminated territories are based on the level of soil contamination and usually on $^{137}$Cs soil contamination. (It is usually suggested that there exists a linear correlation between soil contamination and dose loads on population.) Comparison between disease rates and radiation exposure is usually also based on $^{137}$Cs soil contamination. However, for Belarusian territories with less than 370 kBq/m$^2$ (relatively low contaminated territory) there is no correlation between food and soil radiocesium contamination (i.e. the linear correlation between internal dose and soil contamination does not exist for most Belarussians [Konshin, 1992; Krivoruchko & Makeichik, 1997; Krivoruchko, 1997b]).

Figure 1a presents ordinary kriging estimation of $^{137}$Cs soil contamination for the Belarus territory based on 15,505 measurements in 1992 (Krivoruchko, 1997a; Krivoruchko, Gribov, 1997b). Results of the interpolation are presented as a hill-shaded map and a three-dimensional surface. The arrow indicates the same location for these two surfaces.



a) b)

*Figure 1. a) $^{137}$Cs soil contamination for the Belarus territory, Ci/km$^2$. b) $^{137}$Cs soil contamination near Volozhin City, Bq/m$^2$. Ordinary kriging estimation.*

Figures 1b, 2a, and 2b present the ordinary kriging estimation of $^{137}$Cs soil contamination for three different clusters: the western area between Minsk and Grodno, Figure 1b; the eastern area between Gomel and Mogilev, Figure 2a; and the area between Gomel and Brest on the south of the Republic (this is the area with the highest rates of radiation-related disease in Belarus), Figure 2b. Figure 2c presents the error of ordinary kriging estimation for this area. In this figure, the symbol *x*represents data collection locations (settlements). The darker the color in Figure 2c, the higher the error of estimations. Generally, errors depend on the density of samples and on data variability in the locality. The highest error of estimations is in the southern uninhabited part of this region, which was a former Soviet army testing area.

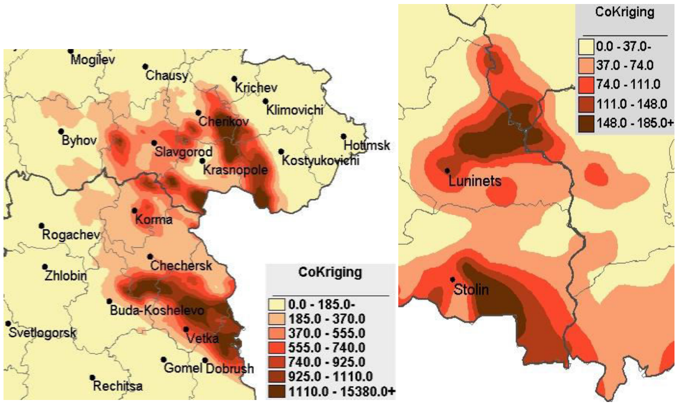
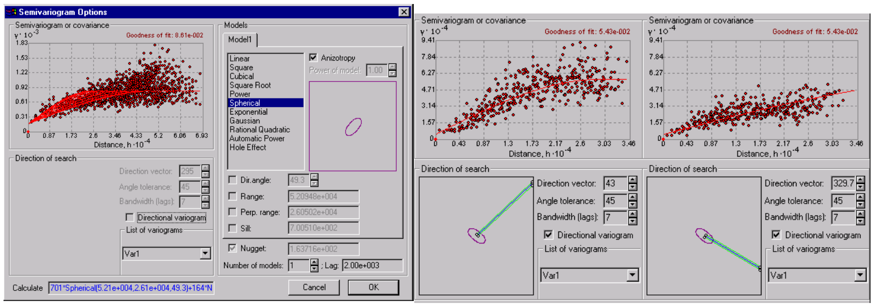


*Figure 2. $^{137}$Cs soil contamination for different parts of Belarus, Bq/m$^2$. a) Eastern area. b) South central area. c) Error map for the south central area. Ordinary kriging estimation.*

*To prepare the maps for the different subregions of Belarus, we used between 500 and 800 samples for each subregion and made kriging consider anisotropy. Figure 3a presents variograms for the western area of radioactivity in Belarus (set of the variogram in various directions) and Figure 3b the eastern area (two directional variograms in perpendicular directions). See the associated maps of contamination in Figures 1b and 2a.*



Goodness of fit: 8.61e-002
Direction vector: 295
Angle tolerance: 45
Bandwidth (lags): 7
☐ Directional variogram
List of variograms: Var1
☑ Anizotropy
Power of model: 1.00
☐ Dir.angle: 49.3
☐ Range: 5.20948e+004
☐ Perp. range: 2.60502e+004
☐ Sill: 7.00510e+002
☑ Nugget: 1.63716e+002
Number of models: 1 ; Lag: 2.00e+003
Calculate 701*Spherical(5.21e+004,2.61e+004,49.3)+164*N

Goodness of fit: 5.43e-002
Direction vector: 43
Angle tolerance: 45
Bandwidth (lags): 7
☑ Directional variogram
List of variograms: Var1

Goodness of fit: 5.43e-002
Direction vector: 329.7
Angle tolerance: 45
Bandwidth (lags): 7
☑ Directional variogram
List of variograms: Var1

a) b)

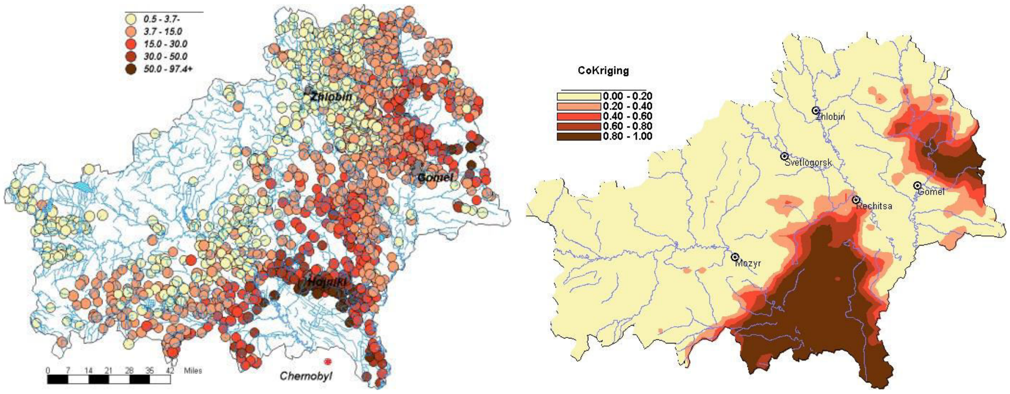
*Figure 3. Directional variograms for $^{137}$Cs soil contamination for two subregions of Belarus: west (a) and northeast (b) areas.*

It is clear that the structure of the contamination depends on the direction and distance from the Chernobyl and it is not appropriate to use standard kriging procedure for the entire data set. To overcome this problem it is possible to use either a so-called stratified kriging (i.e., doing interpolation for different subregions separately and then combining results of the interpolation on a single map) or an approach based on the combination of artificial neural networks and geostatistics to prepare and use a continuous field of correlation between data instead of using unique spatial correlation for the whole area.

The next example of radionuclide soil contamination shows the distribution of $^{90}$Sr in 1994 (kBq/m$^2$) in southeast Belarus, in Gomel province (Krivoruchko, 1997a). The main parameters of the data are as follows: number of samples is 1486, minimum is 0, maximum is 97.3, median is 6.3, mean is 10.7, standard deviation is 12.6, skewness is 2.46, and kurtosis is 10.8.

This data set has an interesting feature. Measurements were not made in the most contaminated areas near Chernobyl and to the east of the Gomel province (see Figures 1 - 3 with data for the whole Belarus territory). People from these areas were evicted. It is important to see how different methods will interpolate the data in such areas.

The map in Figure 4a indicates the locations of sample points by circles 5 km in diameter. We used different deterministic and statistical methods for processing this data set, and parts of the maps were published in (Krivoruchko, 1997a). As was expected, the best results were obtained using different geostatistical approaches. Figure 4b illustrates the successful interpolation of data in the most contaminated areas in spite of missing sample points using probability kriging.



a) b)

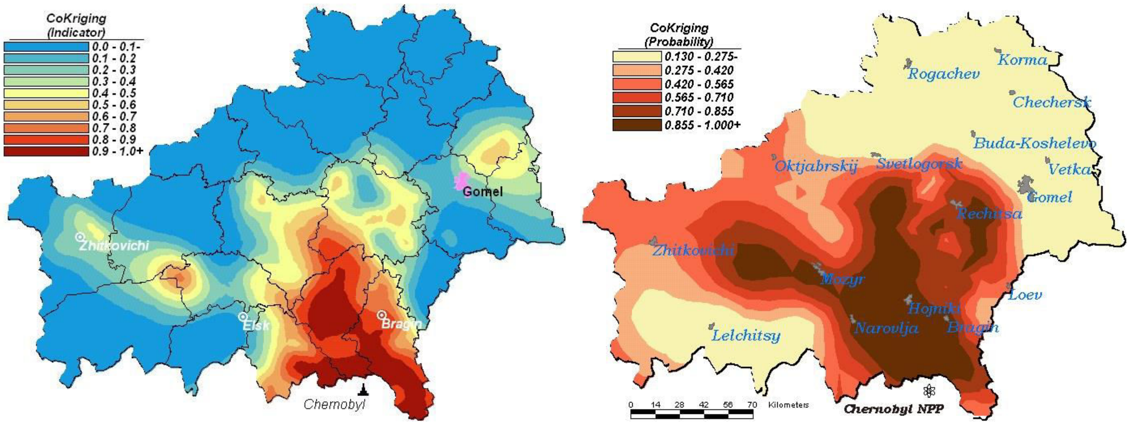
*Figure 4. $^{90}$Sr soil contamination in the Gomel province of Belarus. a) Symbol map. b) Conditional probability that a threshold of 15 kBq per sq.m was exceeded using the probability kriging.*

The geostatistical approach produces reliable results and allows one to know errors of estimations and, as a key advantage, one can produce maps of errors of estimations and maps of probability for a given variable to exceed a chosen threshold. The next example of using this extremely useful method for decision making is illustrated below using plutonium and americium soil contamination data.

Until the middle of the twentieth century plutonium was practically absent in the human. Its concentration in uranium and thorium ores is estimated to be about $10^{-9}$-$10^{-7}$ Bq/g. Substantial amounts of plutonium were released into the natural environment after the beginning of ground nuclear weapons testing in the middle 1960s. In this period, plutonium concentrations in the upper soil layers of Belarus had reached the values of $10^{-5}$-$10^{-4}$ Bq/g. After the accident at the Chernobyl NPP the concentrations of plutonium in soils in the southern areas of Belarus increased to 0.1-0.2 Bq/g. Unfortunately, the biogeochemical activity of plutonium remains poorly investigated.

A peculiarity of the plutonium deposition process after the Chernobyl accident was that the essential part of the radionuclide was associated with fuel particles (hot particles) of different size. This fact can partly explain the existence of outliers and a weaker correlation among the data in some subregions.

We have investigated the spatial structure of  -radiators in (Krivoruchko et al., 1998a; Krivoruchko et al., 1998b) and present some of the results from these articles below. Figure 5a presents the conditional probability that a threshold 370 Bq/m$^2$ of $^{239,240}$Pu soil contamination in the Gomel province in 1992 was exceeded, calculated using indicator kriging. Figure 5b presents the conditional probability that a threshold 100 Bq/m$^2$ of $^{241}$Am soil contamination was exceeded using probability kriging estimation.



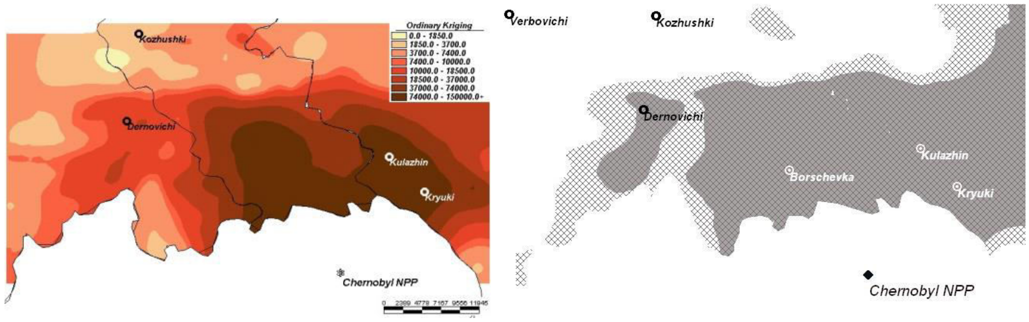
a) b)

*Figure 5. a)* $^{239,240}$*Pu soil contamination in the Gomel province in 1992. Conditional probability that a threshold 370 Bq/m*$^2$*was exceeded using indicator kriging. b)* $^{241}$*Am soil contamination. Conditional probability that a threshold 100 Bq/m*$^2$ *was exceeded in 1992 using probability kriging.*

Because maps of the distribution area of α -radiators have not been published, the common position in the numerous articles devoted to studying the consequences of Chernobyl is to say that dangerous levels of contamination from plutonium are only in the restricted zone where people do not live and that the situation is under control. However, map in Figure 5b indicates that a large number of people are living in the danger zone now: cities, that are represented in Figure 5b are district centers in which population is more than 20,000. Gomel has about 300,000 inhabitants.

The radiation conditions within the restricted 30-kilometer zone around Chernobyl have stabilized. Landscape-related redistribution of radionuclides is completed and pine needles are completely renewed. The population left these territories during the first weeks after the accident. Since that time the restricted zone has been under thorough investigation on various aspects of the problem of radioactive territory contamination. Collecting and processing data on the concentration of long-lived radionuclides in soils is of great importance for deciding on the acceptability of renewing agricultural activity and returning the formerly relocated population. The total amount of α -radiators is changing over time due to the radioactive decay of relatively short-lived radioisotopes of $^{238}$Pu to $^{234}$U and $^{241}$Pu to $^{241}$Am, as well as due to vertical and horizontal migration of isotopes in soil. The migration velocity of α -radiators in the most abundant podzol soils is very small. Over the next hundred years, these radionuclides will remain within the cultivated layer of soil, so we did not take the migration into account.

For the purpose of population dose assessment, it is necessary to account for all α -radiators, having many common features of transfer and retention in tissues. The map in Figure 6a illustrates the predicted distribution of the total amount of α -radiators over Belarus in 2059, at the time of the maximum total concentration of α -radiators.



a) b)

*Figure 6. a) Prediction of the  -radiators in 2059. b) Areas of soil contamination with  -radiators exceeding 740 Bq/sq.m in 1992 (solid pattern) and prediction for 2059 (dashed pattern). Ordinary kriging estimations.*

Figure 6b shows territories with  -radiators surface density of 740 Bq/sq and higher. Solid pattern reflects the contamination in 1992 and the dashed pattern is the prediction for 2059.

The sample area and the total amount of the activity of each radioisotope, as well as the sum of activities of the investigated  -radiators in 1992 and 2059, were estimated using the Monte Carlo method. The sample area was estimated to be 1,180 km$^2$. The total activity of  -radiators in the restricted Chernobyl zone in 2059 will exhibit almost a twofold increase of the activity compared to 1992 and the danger zone will expand. We believe that it will be impossible to live in the restricted 30-kilometer zone around Chernobyl and its immediate borders for the rest of this century and into the next as well. More information about data, variography, correlation between  -radiators and cokriging interpolation can be found in (Krivoruchko et al., 1998a; Krivoruchko et al., 1998b).

## Spatial Distribution of Radiation-Induced Thyroid Cancer

In this section of the article, spatial distribution of radiation-induced thyroid cancer in Belarus is presented based on the results of (Krivoruchko, 1997a; Krivoruchko, Naumov, 1997; Lutsko, Krivoruchko, 1996; Lutsko et al., 1997).

About five years after the Chernobyl nuclear accident, the first suspicions arose about an increased incidence of thyroid cancer in children (Kazakov et al., 1992; Prisyazhiuk et al., 1991). A debate started on whether this increase was real and could be attributed to the release of radiation as a consequence of the accident or whether it was an artifact. In the meantime, with the growing number of cases, the critical voices became silent. The cumulative incidence rate over ten years (1986 - 1995) reached as many as 17.2 cases per 10,000 children in the Bragin district to the east of Chernobyl. It is followed by the Narovlya (16.8 cases per 10,000 children) and the Hoiniki district (12.8 cases per 10,000 children). Last year in Belarus there were about 100 new cases of thyroid cancer in children. Prior to Chernobyl the rate was about one case per the entire Belarus child population.

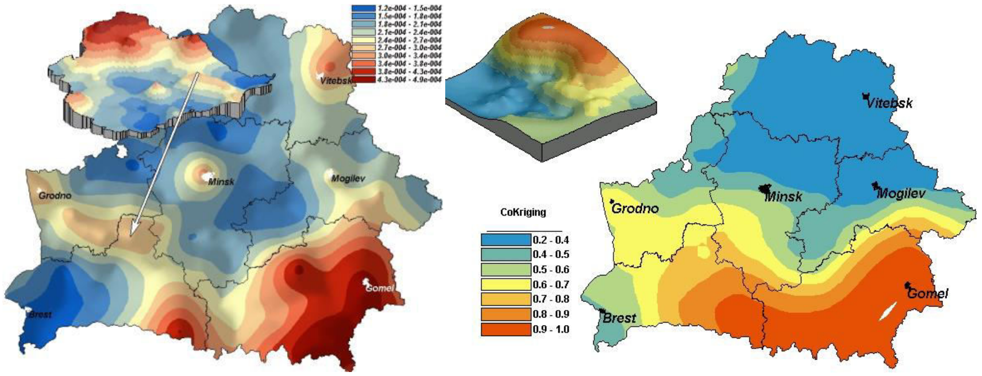
The data on the distribution of thyroid cancer among the population of Belarus was provided by the Minsk Oncology Institute. The database includes 3,773 case records, 902 of which correspond to the pre-Chernobyl period, and the remainder to post-Chernobyl. To define the incidence rate of thyroid cancer in districts and settlements we used information from the 1979 and 1989 censuses of population and information about district population from the annual demographic reports. The age structure of thyroid cancer morbidity in Belarus was investigated in numerous articles (Lutsko et al., 1996).

The cumulative incidence figures for each district were assigned to the coordinates of the district centroids. The 117 incidence values were then considered a spatially distributed variable, that could be interpolated with regard to the spatial autocorrelation found in the variogram.

The feature of the spatial problem of thyroid cancer spatial distribution is that the population strongly differs from one settlement (district) to another. The average number of thyroid cancer cases is two to three per 10,000 over the whole post-Chernobyl period. Thus, the rate of incidence could be correctly defined by the average value at the settlement with population more than 100,000. When population of a settlement is close to 100,000 or less, the rate of incidence becomes inaccurate (there are about 30 districts having less than 30,000 inhabitants). Some cases relate to villages having twenty or more people. When considering the problem of morbidity per settlement one must take into account that uncertainty will rise. In such cases the direct use of the average rate of incidence would be incorrect. One must take into account the initial data structure, which is binomial.

For spatial interpolation of thyroid cancer disease we used a modification of ordinary kriging, which was originally proposed (McNeil, 1991). In order to create a map of the conditional probability that a predefined level is exceeded, we proposed soft indicator kriging modification for binomial data. It excludes from consideration some of the values lying in close vicinity to the threshold value (Krivoruchko, Gribov, 1997a). Thus, in mapping the risk of thyroid cancer development we considered both the distances between settlements or centers of districts and the number of inhabitants. We can say that large settlements with large populations will carry the most accurate characteristics of the epidemic situation in their vicinity, but nevertheless information related to small settlements will not be lost.

Figure 7a represents the map of the estimated risk of thyroid cancer in the post-Chernobyl period (2,756 cases of disease) among all of the population. Figure 7b presents a map of the estimation of thyroid cancer risk for children from 1986 to 1995. The analysis was limited to 528 individuals who were under 15 in 1986. It is presented as the specific probabilities that a predefined threshold of one case of the disease per 10,000 children was exceeded.



a) b)

*Figure 7. a)* The estimated risk of thyroid cancer in the post-Chernobyl period (2,756 cases). Ordinary kriging for binomial data. b) Conditional probabilities that the risk of thyroid cancer in children is higher that one per 10,000. Soft indicator kriging estimation.

Prior to Chernobyl, thyroid cancer incidence was greater in the northern part of Belarus, attaining the highest levels in the Vitebsk district. After the Chernobyl accident the pattern of thyroid cancer incidence substantially changed. Besides the area near Vitebsk, many new areas showed an increased incidence of disease. The territories located along the western trace of the radioactive release suffered the greatest. New areas of increased thyroid cancer risk emerged around Minsk, though iodine doses in this area were relatively low.

There is strong evidence that the increased incidence of childhood thyroid cancer is due to radiation exposure as a result of the Chernobyl accident, based on the geographical and temporal distribution of the cases. The map of the estimated risk of thyroid cancer agrees with the map of the reconstructed surface iodine pollution of the Belarus territory (Krivoruchko, Naumov, 1997; Lutsko et al., 1997) as well as with dose loads on the population in the first days after the accident (see next section).

## Dose Loads on Population in the First Days After the Chernobyl Accident

Due to the unfortunate fact that the measuring equipment was inadequate to properly monitor the scale of radiation exposure during the early period of the accident, detailed direct information on the deposition of the short-lived radionuclides and the doses to the population has been irretrievably lost. Now the only way to reconstruct the dynamics of the radioecologic situation of the initial period of the Chernobyl accident is to make a retrospective assessment of radiation exposures related to the short-lived radionuclides. This can be done by processing extensive empirical information on radiology, meteorology, and epidemiology, and through comparison and validation of the data.

Radioisotopes from the nuclear plant released from April 26 to May 5, 1986, were transferred and deposited over Belarus in a substantially irregular pattern. For the purpose of dose assessment, one must consider the daily structure of radioactive releases in different directions, starting at April 26. For example, over the territories to the west of Pripyat, the former residential community of the Chernobyl nuclear plant, the main clusters of the short-lived radionuclide contamination were formed during the period of April 27 - 30, 1986. To correctly assess the contribution made by short-lived radionuclides to doses received by the population, it is necessary to consider numerous processes, starting with the spectrum of the released activity. To prepare the data for further analysis using GIS and geostatistics, we used 114 exposure dose measurements made on May 10, 1986, and took into account the following (Krivoruchko, Naumov, 1997):

- The passage of the radioactive cloud and analysis of deposition of fuel particles of different natures (aerosols, hot particles, molecular compounds, etc.) in different weather conditions.
- Summarizing the released activity spectrum versus distance (five intervals) and time (average value for the first 15 days after the accident).
- Accounting for correlation between different isotopes as a function of distances (five intervals) and directions (three intervals).
- Extracting 18 radionuclides ($^{239}$Np, $^{99}$Mo, $^{132}$Te, $^{131}$I, $^{140}$Ba, $^{141}$Ce, $^{103}$Ru, $^{89}$Sr, $^{91}$Y, $^{95}$Zr, $^{144}$Ce, $^{106}$Ru, $^{134}$Cs, $^{90}$Sr, $^{137}$Cs, $^{132}$I, $^{95}$Nb, and $^{140}$La) from the experimental dosimetric exposure measurements.
- Establishing dependence between activity in air and soils.
- Estimating the radioactive iodine transfer from grass to cow milk. The thyroid dose can then be estimated from the time-integrated concentrations of radioactive iodine in the thyroid gland taking into account consumption of contaminated milk and leafy vegetables.
- Comparing the calculated and measured doses and soil contamination for different radionuclides.

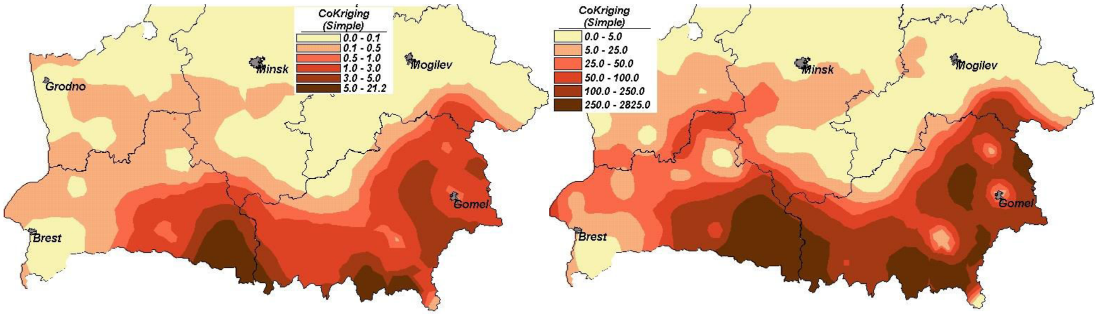
- Comparing the obtained results with epidemiological studies for thyroid cancer in children after the accident.

In areas where no measurements were taken, information was extrapolated and interpolated from the neighboring areas. Such a task was complicated due to the existence of several high peaks of activity near Chernobyl.

Figure 8 presents the results of the reconstruction of the thyroid dose loads from $^{131}I$, based on consumption of contaminated milk by the adult population between April 26, 1986, and May 5, 1986. It was suggested that there was no migration at this time and that people used local foods. Tragically, it is a very reliable suggestion since evacuation from Belarus commenced on May 1, 1986, and by May 5, 1986 only 11,035 people had been relocated.

For the interpolation we used simple cokriging estimation with the normal score transformation of the primary data. The secondary data set was cumulative data of thyroid cancer incidence in children from 1986 to 1995 (117 data samples, one for each of the districts in Belarus).

| **Food** | **UPL, Bq/kg** | **Number of Settlements** | **Number of Measurements** | **Exceeding UPL, %** |
|---|---|---|---|---|



a) b)

*Figure 8. Thyroid dose loads (in mSv) in adult population from April 26, 1986, to May 5, 1986. Simple cokriging estimation. Secondary data set is thyroid cancer incidence in children from 1986 to 1995. a) External gamma doses. b) Internal doses.*

We used ordinary kriging for estimating and mapping the external and internal doses of the population (Krivoruchko, Naumov, 1997). The cokriging approach allows one to reduce the error of estimation when the variables used are related by a common process, as in our case.

| | | | | |
|---|---|---|---|---|
| *Bilberry* | 185 | 317 | 1,383 | 61.03 |
| *Mushrooms* | 370 | 292 | 1,123 | 56.0 |
| *Milk* | 111 | 675 | 19,111 | 14.85 |
| *Pork* | 185 | 229 | 234 | 14.10 |
| *Sour cream* | 111 | 83 | 242 | 12.81 |
| *Well water* | 18.5 | 243 | 2141 | 8.78 |
| *Carrots* | 185 | 252 | 1,439 | 5.84 |
| *Cabbage* | 185 | 182 | 590 | 4.41 |
| *Potatoes* | 370 | 472 | 4,996 | 1.64 |

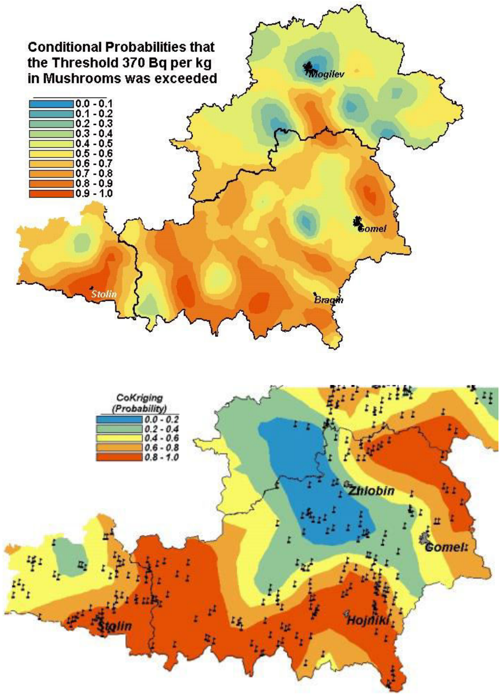
It should be noted that the distribution of short-lived radionuclides is different from long-lived radionuclides (see the section "Long-Lived Radionuclides Soil Contamination") as well as the patterns of dose on population in the very first days of the Chernobyl accident and for 1987-1998 (Krivoruchko, 1998).

### $^{137}$Cs Food Contamination in 1993

Currently, the main sources of radiation hazard to the population living in the contaminated regions are internal exposure from food and external exposure from gamma dose rate in the air. This section is based on the measurements of $^{137}$Cs contents in *83*types of food, that were carried out in the Byelorussian Institute of Radiation Safety in 1993. The cases of $^{137}$Cs content in food exceeding the upper permissible level were published in information bulletins (Nesterenko, 1996), which allows one to identify families under high radiation risk. The initial database containing 53,207 records on $^{137}$Cs concentration in *83* types of food was available for the present investigation. Radiocaesium contamination is distributed very nonuniformly both geographically and within different types of food (see Table I).

*Table I. Exceeding the $^{137}$Cs upper permissible level in some types of food in rural settlements of Mogilev, Gomel, and Brest provinces in 1993.*

For decision making it is important to find regions where it is dangerous to consume food. We have prepared probability maps for each type of food based on the probability kriging approach. Figure 9 presents maps of conditional probabilities that the upper permissible levels for mushrooms and bilberries were exceeded.

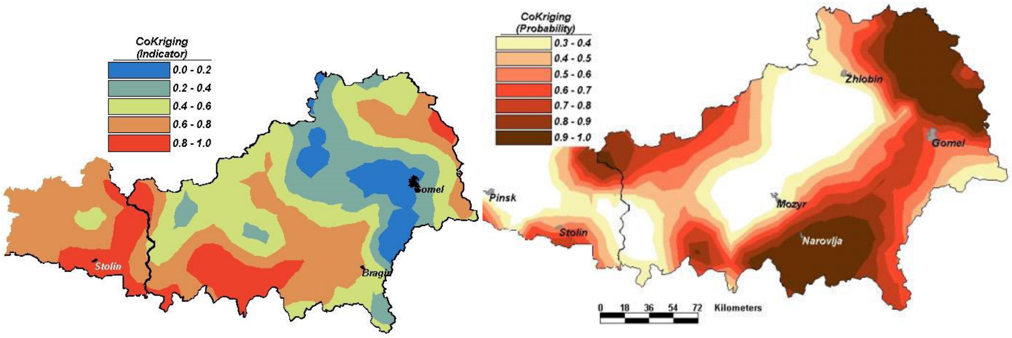


a) b)

*Figure 9. a) $^{137}$Cs concentration in mushrooms. Conditional probabilities that upper permissible level for mushrooms (370 Bq/kg, a) and bilberries (185 Bq/kg, b) were exceeded. Flags indicate the sample locations. Probability kriging estimation.*

A number of factors influence the uptake of radionuclides from soil to plants including the level of soil contamination, the soil type, and the type and extent of countermeasures. We have analyzed the spatial correlation between $^{137}$Cs in the food and in the soil and found that this dependence is very complicated.

$^{137}$Cs milk contamination was changed significantly from 1988 to 1993. Figure 10 presents results of conditional probability estimation that levels of 37 Bq/kg in 1993 and 185 Bq/kg in 1988 were exceeded.



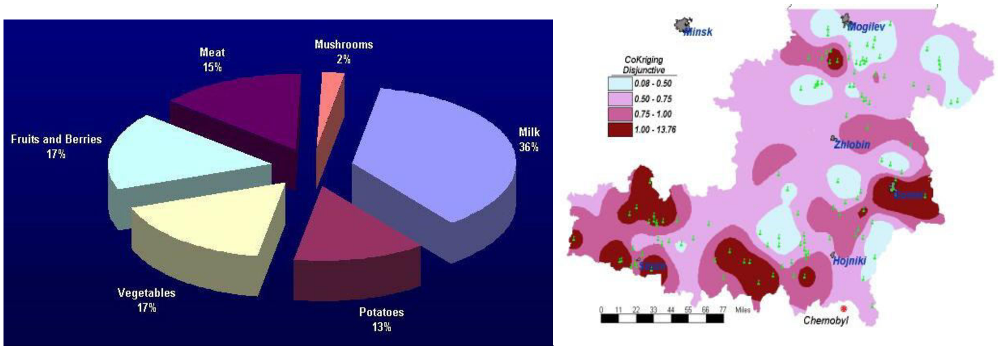
a) b)

*Figure 10. $^{137}$Cs concentration in milk. a) Conditional probability that level 37 Bq/kg was exceeded in 1993. b) Conditional probability that level 185 Bq/kg was exceeded in 1988.*

The maps presented in Figures 9 and 10a are different from the map of $^{137}$Cs soil contamination (Figures 1 and 2), especially in relation to the western passage of the radioactive cloud in the very first days after the Chernobyl accident.

The estimation of the internal exposure was made based on the following food components: milk, potatoes, vegetables, fruits and berries, meat, mushrooms, fish, and bread. We used the dietary habits of the adult rural population of Belarus and information about the diet of children from the Chernobyl zone (Nesterenko, 1996).

Figure 11a presents the structure of the internal dose loads component from food intake for the adult population in 1993, that was based on the information for 120 rural settlements for which we had at least 50 measurements for each food component.



a) b)

*Figure 11. a) The structure of the internal dose loads from intake of $^{137}$Cs with food for the adult population in the 1993. b) Spatial distribution of the internal exposure, mSv per year. Disjunctive kriging estimation.*

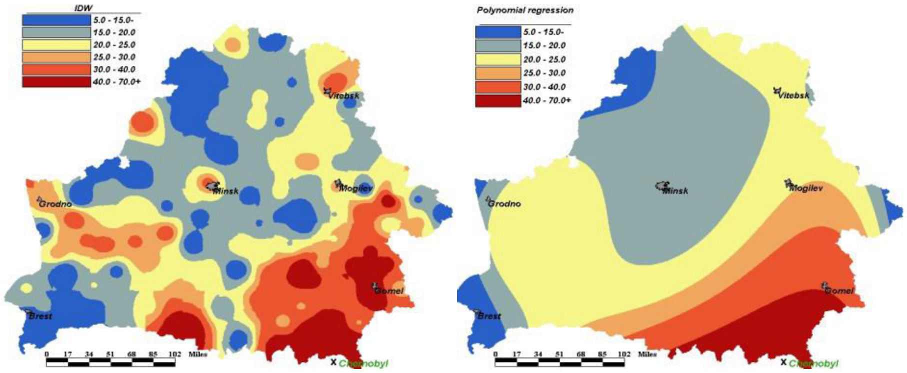
The high radiation risk area covers almost all of the southern part of Belarus. Visual analysis of maps allows one to conclude that the most contaminated food is produced in the rural settlements around the Stolin district, which differs from the official point of view. As was mentioned in (Nesterenko, 1996), practically all children in this area have health related problems. More information about $^{137}$Cs food contamination can be found in (Krivoruchko, 1997b; Krivoruchko, Makeichik, 1997; Krivoruchko, 1998).

## Conclusion

Many methods of spatial data interpolation exist that can analyze data in a reasonable time, however, the choice of a particular method is critical if data is sparse or includes errors. In the section Spatial Distribution of Radiation-Induced Thyroid Cancer, we presented the results of estimating thyroid cancer incidence for the adult population of Belarus after the Chernobyl accident (see Figure 7a). Let us consider the environmental data processing using this example based on some other commonly used methods of spatial data interpolation.

Interpolated values at unvisited points in Figure 12a were obtained by one of the most commonly (due to its simplicity) used method - inverse squared distance weighting. Inverse squared distance weighting interpolation is defined by the following formula $Z^{*}(x_j) = \sum_{i=1}^{N} Z(\mathbf{x}_i) d_{ij}^{-p} / \sum_{i=1}^{N} d_{ij}^{-p}$, where parameter *p* equals 2. The surfaces resulting from a weighted average interpolation depend on the parameter *p* and on the size of a window from which the sample data was drawn. In the case when p⍰ 20, the map will be very close to a Voronoi map. The inverse squared distance weighting approach to data interpolation sends the user in search of the "best" parameters, which, however, cannot be based on sound mathematical considerations.

Figure 12b presents the results of a trend surface analysis technique global polynomial interpolation. This simple deterministic method allows one to produce smooth maps that depend on the order of the polynomial.

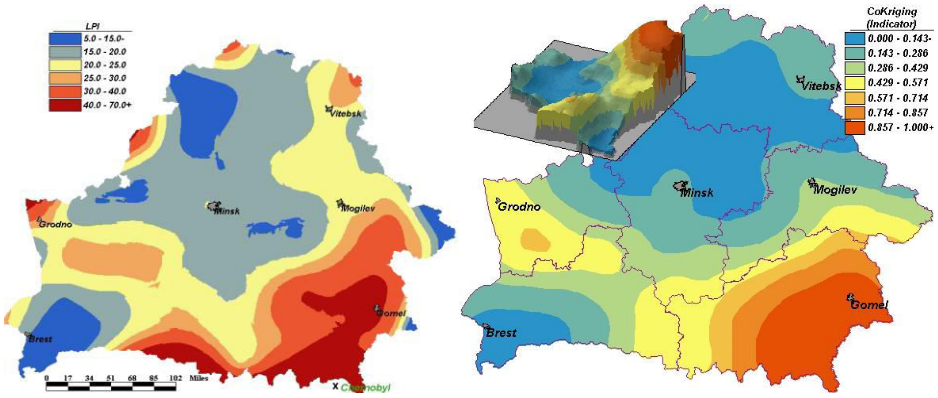


a) b)

*Figure 12. Risk of thyroid cancer in the post-Chernobyl period among all Belarus population. a) Inverse square distance weighting method. b) Global polynomial interpolation. Two-dimensional surface of 5-order.*

Trend surface interpolation is highly sensitive to outliers (extremely high and low values) and it is not an exact interpolator (i.e. resulting isolines do not pass through each of the measurement points).

It is possible to improve this method by using data in the specified neighborhood. Additional improvement of this method can be achieved using the weighting coefficients of the polynomial (Gandin, 1963). Figure 13a displays the results of the local polynomial interpolation by a third-order polynomial. Figure 13b presents the results of standard indicator kriging estimation of the conditional probability that the risk of thyroid cancer is higher than 25 per 100,000. The difference between the map in Figure 13b and the maps in Figures 12 and 13a is apparent.



a) b)

*Figure 13. Risk of thyroid cancer in the post-Chernobyl period among all Belarus population. a) Third-order local polynomial interpolation. b) Conditional probability that risk of thyroid cancer is higher that 25 per 100,000. Indicator kriging estimation.*

The construction of the maps in Figures 12 and 13a raise some important questions, that include:

- What errors would *a priori* suggestions of simple deterministic methods produce? (Is the user even aware of these suggestions?)
- How can one estimate the values of interpolation errors, especially in regions with poor data? How can one take the measurement errors into account?

- What is the probability that the map of interpolation is inaccurate due to uncertainty of the data and errors of interpolation?

To answer these questions we need to use certain advanced spatial analysis methods to understand the errors and uncertainties of the data. Geostatistical techniques provide such possibilities. The advantages of a geostatistical approach to spatial data analysis are the following:

- Problems associated with selecting appropriate parameters for the interpolation are solved using the sound procedures of declustering, normal score transform, detrending, and variogram/covariance modeling.
- It is possible to investigate the anisotropical structure of the data and to use this information for estimation.
- Cokriging allows one to carry out an optimum estimation by taking advantage of the relationships among additional variables.
- Geostatistics allows one to create a map of the probability that a predefined threshold has been exceeded or not.
- Geostatistics allows one to consider measurement errors and to create a map of the error of estimation.
- Using error analysis tools it is possible to find data outliers and to correct their values.
- Geostatistics allows decision makers to make informed decisions about the reliability of maps.

This paper has attempted to justify the use of geostatistical approaches for environmental monitoring and policy planning. The techniques presented can have a profound impact on decision making and policy development. The specific case of Belarus can help us to understand the impacts on the public health and economic vitality of a region greatly affected by a devastating environmental disaster.

## Acknowledgments

I thank Hugh Keegan and Peter Petri for useful discussions.

## References

Deutsch C.V., Journel A.G. (1998) *GSLIB: Geostatistical Software Library and Users Guide*. Oxford University Press, New York.

Gandin L.S. (1963) Translation: *Objective Analysis of Meteorological Fields*. Israel Program for Scientific Translations. Jerusalem, 1965.

*International Basic Standards for Protection against Ionizing Radiation and for the Safety of Radiation Sources*. International Atomic Energy Agency. Vienna (1996).

Journel A.G. (1988) *Nonparametric Geostatistics for Risk and Additional Sampling Assessment*. In Lawrence H.Keith (ed), Principles of Environmental Sampling. American Chemical Society, pp. 45-72.

Kazakov V.S., Demidchik E.P., Astakhova L.N. (1992) *Thyroid Cancer after Chernobyl*. Nature, 359:21.

Konshin O.V. (1992) *Transfer of Cs-137 From Soil to Grass: Analysis of Possible Sources of Uncertainty*. Health Physics, V63, pp. 307-315.

Krivoruchko K. (1997a) *Geostatistical Picturing of Chernobyl Fallout and Estimation of Cancer Risk Among Belarus Population*. Third Joint European Conference and Exhibition on Geographical Information, Vienna, Austria, April 1997, Vol. 2, pp. 676-685.

Krivoruchko K. (1997b) *Spatial Structure of Food Contamination with* $^{137}$*Cs and Estimation of Long-term Internal Dose Loads on Population of Belarus*. International Conference on Low Doses of Ionizing Radiation: Biological Effects and Regulatory Control. Seville, Spain, 17-21 November 1997, pp. 152-155.

Krivoruchko K. (1998) *Radionuclides Soil Contamination, Dose Estimation and Dynamic of Cancer Incidence in Belarus*. Minsk, in print (in Russian).

Krivoruchko K., Gribov A. (1997a) *Estimating the Risk of Cancer in Belarus: Lung Cancer Example*. Ecological Antropology, vol. 2, Minsk-Lublin-Lodz, pp. 423-427.

Krivoruchko K., Gribov A. (1997b) *The Analysis of Spatial Structure and Variability of the Chernobyl Fallouts by Geostatistics*. International Symposium "Actual Problems of Dosimetry", Minsk, Belarus, 28-30 October 1997, pp. 183-194 (in Russian).

Krivoruchko K., Makeichik A. (1997) *Spatial analysis of radiocaesium food contamination in rural settlements of Belarus*. International Symposium "Actual Problems of Dosimetry", Minsk, Belarus, 28-30 October 1997, pp. 195-199.

Krivoruchko K., Gribov A., Figurin I., Karebo S., Pavlushko D., Remesov D., Zhigimont A. (1997) *Mapstudio: The Specialized GIS Integrating Possibilities of Geostatistics*. Third Joint European Conference and Exhibition on Geographical Information, Vienna, Austria, April 1997, pp. 187-196.

Krivoruchko K., Mironov V., Krivoruchko T., Kudrjashov V. (1998a*) About the Plutonium Contamination of the Belarus Territory*. Ecological Anthropology*,* Proceedings of the International Conference: Human Ecology in the Post-Chernobyl Period. Children of Chernobyl, Minsk, pp. 320-323.

Krivoruchko K., Mironov V., Gribov A., Krivoruchko T., Korolev V., Kudrjashov V. *$^{238}$Pu, $^{239,240}$Pu and $^{241}$Am Soil Contamination in the Restricted Chernobyl Zone*. (1998b) Ecological Anthropology, Proceedings of the International Conference: Human Ecology in the Post-Chernobyl Period. Children of Chernobyl, Minsk, pp. 323-325.

Krivoruchko K., Naumov A. (1997) *Reconstruction of dose loads on population in the initial period of the Chernobyl accident and estimation of thyroid cancer risk in Belarus*. International Symposium "Actual Problems of Dosimetry", Minsk, Belarus, 28-30 October 1997, pp. 167-173.

Lutsko A., Krivoruchko K. (1996) *Analysis of methods and reconstruction of iodine doses from the Chernobyl release in Belarus*. International Conference "One Decade After Chernobyl". Vienna, Austria, 8-12 April 1996, Vol. 2.

Lutsko A., Milutin A., Krivoruchko K. (1996) *Health Characteristics of Children and Adults in Belarus, which Suffered from the Chernobyl Accident*. 10 Anni da Chenobyl: Ricerche in Radiologia, Monitoraggio Ambientale e Radioprotezione. Trieste, marzo 1996, pp. 29-38.

Lutsko A., Krivoruchko K., Gribov A. (1997) *Analysis of the Methods and Reconstruction of Iodine Doses from the Chernobyl Release in Belarus*. International Symposium "Actual Problems of Dosimetry", Minsk, Belarus, 28-30 October 1997, pp. 5-12.

Maignan M., Krivoruchko K. (1997) *Integration GIS and Geostatistics: A Software and a Case Study*. Proceedings 18th ICA/ACI International Cartographic Conference. Stockholm, Sweden, 23-27 June 1997, Vol.2, pp. 925-932.

McNeil L. (1991) *Interpolation and smoothing of binomial data for the southern african bird atlas project*. South African Statistical Journal. V.25, pp. 129-136.

Nesterenko V.B. (1996) *Radioecological Monitoring of Food in Rural Settlements in Belarus*. Pravo i economika. N 1-9, Minsk, 1996-1998. In Russian.

Prisyazhiuk A, Pjatak O, Buzanov V.A, Reeves G.K, Beral V. (1991) *Cancer in Ukraine, post-Chernobyl*. Lancet, 338:1334-5.

Rivoirard J. (1994) *Introduction to Disjunctive Kriging and Non-Linear Geostatistics*. Oxford Univ. Press. New York.

Sakharov A. (1958) *Radioactive Carbon from Nuclear Explosions and Nontreshold Biological Effects*. Soviet Journal of Atomic Energy, vol.4, N6.

**Konstantin Krivoruchko**
Environmental Systems Research Institute, Redlands, California

## #2. Linking GIS and Geostatistics.

As evidenced by current pop music, just about anybody can create orchestral compositions using modern electronic keyboard equipment without having mastery or even experience with individual instruments. Is it likewise possible to embed modern geostatistical software inside a Geographic Information System (GIS), which can be used without an exhaustive in-depth understanding of geostatistics?

Geostatistics are statistical methods to describe spatial relationships among sample data and to apply this analysis to the prediction of spatial and temporal phenomena. In addition to the description of spatial patterns and interpolation of data, important components of geostatistical analysis include error analysis and the integration of secondary data in prediction algorithms through the use of cokriging. One can assess the performance of the interpolation as well as detect errors in the source data using error analysis, which includes mapping the error of estimation and performing both validation and cross-validation. Cokriging allows one to perform an optimum estimation while considering more than one variable. The spatial relationships among several associated variables (easily captured and manipulated with GIS technology) may be used to improve the estimation of another variable, which is difficult to measure, by using data, which can be collected more easily.

Sampling and mapping in the earth sciences are complicated by complex spatial and temporal variations. The structure and intensity of patterns being sampled often cannot be determined or predicted reliably with deterministic models because of uncertainties in the data and the phenomena under investigation. The best we can do using interpolation and estimation methods is to be as objective as possible and to consider the interrelations of the data under investigation.

Deterministic approaches to interpolation (trend surface, inverse distance weighting, triangulation, and splining) are based upon a priori mathematical models of spatial variation. They assume the sampled data has no errors, which is often an incorrect assumption. In practice, error can not be eliminated but only minimized. Therefore, in most cases one cannot produce the best representative map of estimated values in unsampled locations with these techniques.

Currently, in order to integrate advanced methods of spatial data interpolation including geostatistics into GIS, users are required to use separate statistical packages to process and store the results of interpolation into a GIS supported format. To analyze complicated environmental problems and to present the results of an analysis on a map requires a group of specialists who are familiar with both GIS techniques and complicated statistical software. In general, statistical software, which includes

geostatistical tools, can be divided into two groups:
1. Simple programs which include standard ordinary kriging estimation along with a few additional tools.
2. Advanced programs with a complete compliment of geostatistical tools for any type of data.

Using programs from the first group does not promote an understanding of the advantages of geostatistics. The time burden of mastering software from the second group has been an enormous impediment for its routine use by researchers and students, who are utilizing GIS technology. As a result there are many articles devoted to spatial interpolation in agriculture, meteorology, environment and other disciplines where geostatistics are used inaccurately- primarily because researchers used non-optimal data processing techniques.

We shall discuss an approach of combining advanced spatial statistics and GIS. This combination was first presented in the software package, GIS MapStudio [1-3], which integrated both modern GIS visualization techniques and easy to use geostatistical methods of spatial data analysis. These methods are currently being considered for further development and implementation into ESRI’s GIS product line.

The advantages of a geostatistical approach to spatial data analysis are:

- Problems associated with selecting appropriate parameters for the interpolation are solved using the sound procedures of declustering, normal score transform, detrending and variogram/covariance modeling.
- It is possible to investigate the anisotropical structure of the data and to use this information for estimation.
- Cokriging allows one to carry out an optimum estimation by taking advantage of the relationships among additional variables.
- Geostatistics allows one to create a map of the probability that a predefined threshold has been exceeded or not.
- Geostatistics allows one to consider measurement errors and to create a map of the error of estimation.
- Using error analysis tools it is possible to find data outliers and to correct their values.
- Geostatistics allows decision-makers to make informed decision about the reliability of maps.

An implementation of such tools in a commercial GIS package would provide a user-friendly interface for the rapid analysis and display of data using geostatistical tools within a GIS environment without the requirement of an in-depth understanding of the statistical techniques for:

- Spatial correlation analysis and modeling of multivariate data sets.

- Creation of continuous surfaces from sampled locations.
- Error analysis and quantifying uncertainty for data sampling.
- High-quality visualization of the interpolation results.

Examples of data analysis and spatial interpolation will be presented and then discussed. The answer to the question about the possibility of implementing sound geostatistical analysis within a GIS will be known by ESRI product users in the near future.

**References**

1. Krivoruchko K., Gribov A., Figurin I., Karebo S., Pavlushko D., Remesov D., Zhigimont A. Mapstudio: The Specialized GIS Integrating Possibilities of Geostatistics. Third Joint European Conference and Exhibition on Geographical Information, Vienna, Austria, April, 1997, pp. 187-196.
2. Krivoruchko K. Geostatistical Picturing of Chernobyl Fallout and Estimation of Cancer Risk Among Belarus Population. Third Joint European Conference and Exhibition on Geographical Information, Vienna, Austria, April 1997, Vol. 2, pp. 676-685.
3. Krivoruchko K., Maignan M. Integration GIS and Geostatistics: A Software and a Case Study. Proceedings 18th ICA/ACI International Cartographic Conference. Stockholm, Sweden, 23-27 June 1997, Vol.2, pp. 925-932.

**Mei Po Kwan**
Department of Geography, The Ohio State University, Columbus

## Analysis of Human Spatial Behavior Within GIS: Recent Developments and Future Prospects

In the past few years, tools for performing spatial analysis in a GIS environment have become more available to researchers. For instance, functionalities of both S+SpatialStats and SpaceStats are now accessible through the more user-friendly interface of ArcView GIS. Along with this development, comprehensive digital data sets of metropolitan areas collected and maintained by public agencies are also becoming widely available. As these geographic data sets often contain detailed information of urban areas (e.g. attributes of all land parcels and the transportation system) not available before, they render the application of new analytical methods possible, especially when these methods required the data handling capabilities of GIS. Lastly, the increasing availability of individual-level geo-referenced data (e.g. geocoded travel diary data) also makes individual-based spatial analysis feasible. As a result of these three developments, new opportunities for the analysis and theoretical understanding of disaggregate human spatial behavior are emerging. The most important of these pertains to the representation of the objective and subjective environment within GIS, as well as the possibility of person-based and frame independent spatial analysis. These future prospects are briefly outlined as follows.

First, recent trends in the availability of analytical tools and data allow for the realistic representation of the complex objective environment for the analysis of human spatial behavior (Kwan 1997). If detailed attributes of land parcels and the transportation systems can be represented through incorporating the relevant information into a comprehensive geographic database, the analyst may go beyond the simplified and geometric operationalization of geographic constructs as often done in traditional spatial analysis. For example, instead of using the straight-line distance between two locations, the actual travel distance over the transportation network can be used (as in Talen 1997 and Kwan 1998). Further, given the more realistic geographic environment represented in the GIS, it is possible for the analyst to perform non-isotropic spatial analysis, which does not depend on any assumed spatial distribution of opportunities in the urban environment (Tobler 1993).

Second, with appropriate data collection effort and using the spatial data handling capabilities of GIS, elements of individual cognitive map which bear upon spatial behavior may be incorporated into analytical models (Golledge et al 1994; Kwan and Hong 1998). By taking into account factors which affect human spatial behavior (e.g. cognitive and space-time constraints) through establishing more realistic representations of the subjective environment, spatial analysis in a GIS environment can be based upon the more relevant effective environment of individuals. This will extend the theoretical foundation of spatial analysis to include the behavioral dimensions into the analytical framework (Fotheringham 1993).

Third, using geo-referenced individual-level data in a GIS, spatial analysis will no longer be affected by any prior zonal or areal partition of the study area (as in the case where socio-demographic data are aggregated based on a zonal schema) (Kwan 1998). This implies a shift from traditional methods to new techniques for specific problems. For example, point-pattern techniques (such as cross K-function) may be more appropriate than conventional zone-based methods for measuring individual accessibility to urban opportunities when individual-level data are used. This, in other words, allows for the use and development of 'frame independent' spatial analytical methods (Tobler 1989), which may help ameliorate the modifiable areal unit problem.

Changes in the above three areas will allow for the application of new methods to specific problem areas pertaining to human spatial behavior. Further, by placing the individual into the focus of spatial analysis, and with considerations of both the objective and subjective environment, such person-based and frame independent framework will enable the examination of fine-scaled, inter-personal differences based on gender, race, or other socially significant categories. This, perhaps, could be the beginning point for a mode of spatial analysis which is more congenial to poststructuralist and feminist conception of space and the individual. Obviously, much research is needed to examine this possibility,

**References**

Fotheringham, A.S. (1993) "On the future of spatial analysis: The role of GIS" Environment and Planning A 25:30-34.

Golledge, R.G., M.-P. Kwan and T. Garling (1994) Computational-process modeling of household travel decisions using a geographical information system. Papers in Regional Science, 73(2):99-117.

Kwan, M.-P. (1997) GISICAS: An activity-based travel decision support system using a GIS-interfaced computational-process model. In Activity-Based Approaches To Travel Analysis, 263-282, edited by D.F. Ettema and H.J.P. Timmermans. Pergamon: New York.

Kwan, M.-P. (1998) Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. Geographical Analysis 30(3): 191-216.

Kwan, M.-P. and X.-D. Hong (1998) Network-based constraints-oriented choice set formation using GIS. Geographical Systems, in press.

Talen, E. (1997). The social equity of urban service distribution: An exploration of park access in Pueblo, Colorado, and Macon, Georgia. Urban Geography 18:521-541.

Tobler, W.R. (1989). Frame independent spatial analysis. In The Accuracy of Spatial Databases, 115-122, edited by M. Goodchild and S. Gopal. Taylor and Francis: London.

Tobler, W.R. (1993) Three Presentations on Geographical Analysis and Modeling. NCGIS Technical Report 93-1.

**Tony Lea**
Compusearch Micromarketing Data and Systems

## GIS and The Promotion of Geographical Analysis in Business - A Pessimistic View From The Trenches

The following observations and thoughts relate to the use of spatial analysis and GIS in business, especially the marketing and retail site location dimensions of business. To assist in understanding the context of the remarks, here are a couple of sentences of background information. I am a former academic who has taught both spatial analysis and GIS (University of Minnesota, University of Toronto, Ryerson Polytechnical University). I have spent many years (since 1980) working as a consultant in the area of market analysis with a strong geographic or spatial component. I am currently senior vice president at Compusearch Micromarketing Data and Systems which has recently become a division of the Polk Company (Detroit). Compusearch provides a wide range of data, software, project work and consulting in the area that has become known as Business Geographics. The majority of our clients are private firms in almost every sector including retailing, financial services, automotive, telecommunications, packaged goods, media, etc. The data we sell includes a wide range of census and administrative area boundary files, census data, street and highway files, estimates and projections, automotive and track registration data, point files including facilities such as shopping centres, department and mass merchandise stores, grocery stores, etc. Our premier software products are user friendly business/marketing analysis software systems with all relevant data sets integrated in turn-key systems. The packages are very geographical by design and feature customized (ESRI) ArcView applications or (ESRI) Map Objects as the geographical interface.

There has been very significant progress in the last 10 years or so in designing and producing user friendly GIS oriented software with huge amounts of diverse data directly accessible to marketing and business users. Five years ago the most powerful of such systems were Compass by Claritas, “Infomark” by Equifax - National Decision Systems and “Conquest” by Donnelly Marketing (all of the U.S.) and “Mosaic Systems” by CCN Marketing (Nottingham, U.K.). In addition, at that time, thousands of businesses of all kinds used one of the following less expensive (and less integrated) GIS software packages. MapInfo, Atlas GIS, Tactician, Transcad, SPANS, Maptitude, SCAN US and several other small shape packages

The first mentioned 4 integrated packages made geographical analysis and mapping very close to push-button and were designed for relatively unsophisticated and non-quantitative users like marketing managers and real estate researchers. On the other hand, the various general-purpose GIS desktop systems were viewed as lower end hands-on type of systems demanding an in-house analyst (typically a geographer) by the business community. Data had to be purchased separately and often took a long time to get set up for easy use by the software. In general, however, the standard GIS packages were very slow especially on larger (national and state scale) data intensive applications.

Up to about 5 years ago both the integrated geodemographic/GIS packages and the general purpose desktop mapping software systems were used to provide answers to very simple questions and the methodology was almost always very simple. Some typical applications follow:

1. Geocode customer addresses
2. Map customers in a market as dots
3. Map customers, or customer dollars, over household penetration as small area choropleth maps
4. Create circular trade area around a store and extract absolute and relative profiles of the people and households who live there
5. Create a geodemographic profile of customers based on weighted census variables or membership in geodemographic clusters
6. Create choropleth maps of all the block groups in a market based on extent of matching to the profile in 5

Most of these tasks barely qualify as analysis let alone geographical analysis.

How have things changed in terms of business users objectives, tools and analysis in 1998. Regrettably, very little in terms of analysis.

First the desktop GIS software field has narrowed considerably. A new entrant ArcView from ESRI quickly gained substantial market share. Now MapInfo and ArcView likely have 90-95% of the business market for desktop GIS software, outside of the higher end market for the integrated type of geodemographic/GIS systems (referred to above). Several of the other software packages have virtually disappeared from businesses AtlasGIS and SPANS as the GIS software industry consolidates. The other small share packages have lost share. There is now less choice in low-end desktop software. The top dominant packages have added very little new functionality especially in terms of analysis.

There has been a very significant restructuring in the American geodemographics and business GIS industry. First Strategic Mapping bought Donnelly Marketing and discontinued Conquest. Then ESRI bought Strategic Mapping, kept the Atlas software line and sold the data business to Claritas. More recently, last year, Claritas bought Equifax-NDS, its long time arch rival. It is very likely that Claritas will discontinue the Infomark geodemographic data and software package. Claritas has reworked its Compass product with a MapInfo-based GIS engine and given it a new name.

There has clearly been a substantial reduction in competition and in choice for users in this industry in the U.S. The general philosophy of the few big remaining players is to try to focus even more on the off-the-shelf mass market and add only very simple (but usually sexy) new functionality for business users. The number of buttons in the software has increased but the sophistication has not.

There are likely two exceptions to this general trend and both involve vertical applications. Lower end sales territory optimization and truck/bus routing applications have become more sophisticated. Both these applications now permit some interesting analyses at a reasonable price. But, perhaps the highest demand applications-relating to retail site evaluation and sales forecasting continue to be overly simplistic or even crude when evaluated by the standards of spatial analysis. A few specialized consulting firms offer services to build spatial interaction models market by market for interested financial institutions (especially banks) and large retailers, but in general these approaches are viewed as too expensive and perhaps over-kill.

It is very clear that the software and data vendors of this industry view most business GIS users as:

1. Extremely price sensitive
2. Rather unsophisticated (keep it simple, stupid)
3. Quite non-analytical
4. Wanting simple answers quickly without much concern for the quality of the numbers
5. Still titillated by colour graphics and maps
6. Unwilling to become involved in new R&D processes

There is a good chance that the software designers and data vendors are right.

I have been involved in putting forward many research proposals to major North American firms to build high quality GIS-based systems for their distinctive business problems. It is a very hard sell. In general, business users really do want to keep things simple and avoid esoteric methodological issues.

I have mixed views of how things could trend in the future. I think that it is possible for a group of exciting and competent geographers to get at least some business users excited about doing things right or more rigorously. However, with the number of charlatans around praying on the business people who just want a number quickly, and with the general price sensitivity of businesses, I think it is really unlikely to happen. On the other hand things could be worse - with simpler software and even cruder ‘analysis tools’. Business users could demand even lower priced and quicker answers to their questions.
There has been some calls in papers in the trade journals for this type of ‘progress’.

For those of us who are delighted to become involved in genuine spatial analysis to derive the highest quality research for our business and related clients the remarks

above have the following implications:
1. High-end spatial analysis research for the business community (with a few exceptions) is done by dedicated researchers without the full knowledge and support of the clients since they are not being paid for all they are doing for the client.
2. Clients never seem to pay full price for this high end research much of the real quality analysis is thrown in gratuitously.
3. Most of the analysis work is done in software outside of GIS proper either in statistical packages or in database packages of in 3 or 4 GL programming languages.
4. The work that is done in GIS is high end GIS functionality and this seems inevitably to be done (until lately) in ArcInfo.
5. Much time is spend moving data between the various software packages that seem to be best for different purposes
6. After great analyses are completed, analysts have to spend often more time than the analyses took on figuring out how to present the methods and findings as "essentially simple", to please the executives who write the cheques..

I believe that there are few forces capable of intervening in this state of affairs which will effectively promote higher quality analyses and models to businesses at least on a scale that is large enough to make a really noticeable dint.

Final Remarks.

I have asked a number of my colleagues at Compusearch, in academia, and in industry to make some comments on this draft and to try to prove my apparent cynicism wrong.

I suspect that I will receive some good feedback from this document and would be happy to update this paper with new views in the near future.

**Micahel Leitner**
Department of Geography and Anthropology, Louisiana State University, Baton Rouge

Recent years in North America have seen a rapid development in the area of crime analysis and mapping using Geographic Information Systems (GIS) technology. In 1996, the US National Institute of Justice (NIJ) established the crime mapping research center (CMRC), to promote research, evaluation, development, and dissemination of GIS technology. Since 1997, the CMRC has been organizing annual symposia on crime mapping. In addition, a CRIMEMAP listserv, designed to get crime analysts, researchers, geographers, and other interested parties communicating about computerized mapping technologies related to criminal justice applications has been established, as has a Visiting Fellowship Program to support the research and development of topics in the area of crime mapping and the spatial analysis of crime.

The long-term goal of these and related activities is to develop a fully functional Crime Analysis System (CAS) with standardized data collection and reporting mechanisms, tools for spatial and temporal analysis, visualization of data and much more. Among the drawbacks of current crime analysis systems is their lack of tools for spatial analysis. For this reason, spatial analysts should research which current analysis techniques (or variations of such techniques) that have been already successfully applied to other areas (e.g., epidemiology, location-allocation analysis, etc.) can also be employed to the spatial analysis of crime data.

The following lists current problems that hamper the development and availability of crime analysis systems in the law enforcement community. I believe that these issues are not only restricted to crime analysis systems, but are also applicable to other areas of GIS and spatial analysis applications.

1. Money and time constraints

Police departments have limited resources to purchase computer hardware and GIS, spatial analysis, and mapping software. Training police officers with this new technology costs further money and time. Building a comprehensive crime database that can be implemented, shared, and updated among the different units in a police department costs additional time and money. Smaller police departments (less than 1000 employees), because of very limited financial resources, are especially handicapped to fully participate in this new development. On the other hand, larger police departments have already established their own crime analysis units including full time personnel to apply for research grant money (examples are New York City, Chicago, Washington D.C., Los Angeles, etc.)

2. Confidentiality, security, and accessibility of crime data

Crime data are originally collected at the individual level. Police reports usually record individual data (sex, age, race, address, etc.) on the victim, the offender (if apprehended), the crime location, and time of incidence. This involves important issues of data confidentiality, security, and accessibility.

3. Lack of training in GIS, spatial analysis, and computer mapping

Police officers and detectives lack training in these new technologies and might even be computer illiterate. For this reason any software product developed for law enforcement agencies needs to be easy to use, preferably with a point and click interface, steep learning curve, and appropriate default values. It is no surprise then that MapInfo and ArcView are the currently two most often used GISs among police departments in this country. Additionally, the CMRC has coordinated six different modules on crime mapping, the use of GIS, and spatial analysis that will be taught to interested law enforcement agencies nationwide.

4. Lack of spatial statistical software that is targeted specifically for crime analysis

The development of spatial crime analysis tools and their implementation into existing commercial GIS packages requires the combined effort of law enforcement agencies, academia, and software engineers. The perhaps first stand-alone crime analysis package is the so-called Spatial and Temporal Analysis of Crime (STAC) package, which was developed by the Illinois Criminal Justice Information Authority in the mid-80s. This software possesses various spatial and temporal analysis tools to detect patterns of crime in a community, using both geographic and time data (Bates 1987). The spatial analysis part is restricted to calculating the nearest neighbor index and the identification of crime hot spots. A second example stand-alone analysis package is POINTSTAT, a spatial statistics program for the analysis of point locations which can be used to analyze events (e.g., crime incidents, accident locations) or the spatial distribution of particular types of organizations (Levine et al. 1994). The most recent example is the so-called Point Pattern Analysis (PPA) software, developed by Chen and Getis (1998). A major drawback of these programs is that they are purely analysis tools and need to be loosely coupled with a GIS (e.g. Arc/Info, MapInfo) for displaying the analysis results (e.g., crime hot spots).

An example of an extension kit to an existing GIS is CrimeView by the Omega Group, a suite of integrated crime analysis tools designed for use in the object oriented ArcView GIS environment. Such crime analysis systems provide the full functionality of a GIS tailored to the specific needs of law enforcement agencies (e.g. standardized data collection and reporting mechanisms; tools for spatial and temporal analysis as well as visualization) and to the specific nature of crime data (i.e., point location in time).

Projects currently underway include the development of new GIS procedures for analyzing incident data (Levine and Wong 1997), for predictive modeling and for enhancing proactive policing (Hunt and Zubrow 1997). Current spatial point pattern analysis software that has been already applied to geographical epidemiology is summarized in Gatrell et al. (1996).

I am very interested to participate in the workshop on status and trends in spatial analysis. This workshop provides me with the opportunity to find out which new spatial analysis tools already applied to other areas of geographical inquiry might also be applicable to the spatial analysis of crime. It also gives me the opportunity to share with the workshop participants recent trends in spatial crime analysis, associated needs, and problems.

Together with one of my graduate students, I am currently developing computer programs that calculate the spatial association between two or more point data sets (e.g., crime locations, offenders and victims residences, etc.). Such techniques are applicable to a wide array of geographical problems. Another one of my graduate students currently develops models of spatial crime displacement due to urban renewal programs that have been carried out in the city of Baltimore, MD. Such models will help city planning bureaus and law enforcement agencies with pro-active decision making with the goal to decrease criminal activities.

**References**

Bates, S. 1987. Spatial and temporal analysis of crime. Chicago: Research Bulletin, Illinois Criminal Justice Information Authority.

Chen, DongMei, and A. Getis 1998. Spatial Point Pattern Analysis (PPA) Software. Department of Geography, San Diego State University.

Gatrell, A.C., Bailey, T.C., Diggle, P.J., and B.S. Rowlingson 1996. Spatial point pattern analysis and its application in geographical epidemiology. Trans Inst Br Geogr, NS 21:256-274.

Hunt, E.D., and E.B.W. Zubrow 1997. Building Crime Analysis Extensions for ArcView. Arc/Info User Conference. CA

Levine, N., Khoo, C., Okazaki, D., Kim, K., and L. Nitz 1994. Pointstat: A statistical program for analyzing the spatial distribution of point locations (version 1.0). Honolulu, HI: University of Hawaii at Manoa.

**Yee Leung**
Department of Geography,
Center for Environmental Studies, and Joint Laboratory for GeoInformation Science
The Chinese University of Hong Kong

GISs, by their inception, are spatial data processing systems with efficient mechanisms for data storage, retrieval, and display. Though it has been claimed that GISs are capable of performing spatial analysis, they, especially canonical GISs, are in fact of very limited capability in this aspect. The kinds of analysis GISs can offer are rudimentary data manipulation procedures ranging from overlay and buffer to simple statistical analysis. They are by no means a system that can entertain complicated spatial analysis tasks. Though developments in recent years have greatly appended into GISs more advanced spatial analysis tools, their capabilities in spatial analysis and decision making are still far from satisfactory.

Due to the complexity of issues in spatial analysis, it is actually unrealistic to ask for a GIS to include in its functionalities all aspects of spatial analysis. GISs, after all, are just data processing systems which, in my view, should just stay as a data processing system for spatial data storage, retrieval, and display. We should not ask a GIS to lead our analysis. On the contrary, GISs should be used as a support to facilitate spatial analysis and decision-making. We should separate the two things but provide an efficient and user-friendly environment for their integrative utilization.

To achieve such a goal, we need to build GISs as a truly open system with which we can customize for specific spatial analysis in an efficient and effective manner. Closed systems are doing a disservice to the accomplishment of such an objective and will perpetuate the current state of the use dictated by most of the commercial products, i.e. merely a device for data processing and display with limited spatial analytical capability. Though some spatial analysis modules have been incorporated into some GIS products, and macro languages have been provided for customization, they can hardly be considered as open systems in the strict sense.

The open-system design should pay particular attention to the entertainment of the following major movements in spatial analysis:

1. Spatial Dynamics. The analysis of dynamics in spatial structures and processes have unique requirement of data structure, I/O, and data-model integration.The concept of temporal GIS has been around for quite some time, and yet a truly spatio-temporal GIS is still at large.

2. Evolutionary Computation. Spatial analysis in recent years has experienced an upsurge in the use of two fast evolving paradigms: Evolutionary and neural computation for complex systems analysis. The design of GIS suitable for the requirement of models

such as neural networks, genetic algorithms, and evolutionary programming is of importance.

3. Artificial Intelligence. The availability of intelligent SDSS is of practical value to researchers and practicing professionals. The design of GISs for an effective support of AI oriented investigations is necessary.

4. Uncertainty. Though the issues of uncertainty in GISs have been investigated over the years, a product that can truly convey uncertainty in GIS operations and spatial analysis is still non-existent. This is a totally unacceptable state of the art provided by commercial products.

As a concluding general observation, if one takes a careful examination of GIS related researches/publications, it is apparent that quite a large number of them can be accomplished without a GIS. They are using a modern means to achieve an old task with very little contribution to the advancement of spatial analysis. If this situation perpetuates, GIS will run out of gas in the near future and GIS research may boil down to nothing but the use of commercial software to doctor up our analysis, or at most to make data management and display more efficient. To make GIS research sustainable, and to further develop the discipline, we have to look for an answer in spatial analysis. After all, we want to solve spatial problems with GISs. Therefore, issues involving the design and integration of GISs with spatial problems such as those discussed above will give us a more promising future. Of course, doing all these in the internet should also be in the agenda.

**Paul Longley**
Department of Geography, University of Bristol

## 1. Introductory comments

'Econometric theory is like an exquisitely balanced French recipe, spelling out precisely with how many turns to mix the sauce, how many carats of spice to add, and for how many seconds to bake the mixture at exactly 474 degrees of temperature. But when the statistical cook turns to the raw materials, he finds that hearts of cactus fruits are unavailable, so he substitutes cantaloupe; where the recipe calls for vermicelli he uses shredded wheat; and he substitutes green garment dye for curry, ping pong balls for turtle's eggs, and, for Chalifougnac vintage 1883, a can of turpentine.' Valavanis 1959: 83, quoted in Kennedy 1979.

Ever since undergraduate days in Bristol in the 1970s, I have felt fully imbued with the quantitative locational analysis tradition in geography - not least because some of the origins to the approach can be traced to Bristol in the 1960s. Yet I share the frustration aired in some of the other position papers that the spatial 'mainstream' to geography has been sidelined in the major geography journals, that it accounts for a reduced real share of intellectual activity in the subject, that its interdisciplinary outreach has been limited, and that today's GIS practice appears to develop largely separately from academia. (At least GIS and RS together make up one of six 'specialisms' that are key to UK central government's ranking of subject performance in what remains a mainstream discipline). I should like to comment on the way that research practice and, in particular, data handling, may contribute to this state-of-affairs, and to suggest how reconfiguring some priorities in spatial analysis might be beneficial.

Spatial analysis, like econometrics, has benefited from the proliferation of digital data sources in recent years. Today's spatial data models allow far 'thicker' depictions of geographical reality to be created than those I cut my own teeth on. The transformational (Martin 1996) or simplifying assumptions entailed in building GIS-based models of real world spatial distributions have become much less heroic as a consequence. And of course it is well known that developments in computer hardware remain more or less commensurate with the increase in available data, making it possible to explore and model spatial interactions in more detail than ever before. Viewed in this context it is paradoxical that, in the UK at least, there is less faith than ever in 'predict and provide' approaches to planning, that business and service planning is turning away from conventional spatial analysis and that some of the spatial analysis community view essentially 'black box' techniques with increasing favour. Why is this the case?

In the socio-economic realm one suggestion might be that any increase in the sophistication of analytical models has been more than outpaced by increases in the complexity of the systems themselves - witness, for example, the scale and pace of change in the physical forms of urban systems, or the fragmentation of household consumption patterns and lifestyles. A variant on this theme is suggested by Curry (1995) and others who seem to suggest that the quality of digital data can never be adequate for the resolution of significant problems of real world concern. A third suggestion, which is the one I will pursue here, is that the research community should refocus effort away from abstract semantic discussion or analytical elegance and towards the messy empirical problems of data integration. This should be done in as rational, orderly and application-centric a fashion as possible.

Goodchild and Longley (1999) appraise of the 'linear project design' as a model for contemporary research in natural and social science. For generations of students, the formulation of research hypotheses has been followed by choice of a data collection method (and designing a survey schedule, as appropriate), identifying a sample design, piloting, field collection of data (with verification and resampling), collation of results, analysis, and report-writing. They reflect that, although this robust and defensible schema has underlain generations of student dissertations, it was never a panacea in practice, for reasons of data resolution, surrogacy and timeliness - and the amount of funding available for scientific research (we're all researchers now!). Today's GIS environment is also characterised by datasets which are collected by many different means and which pass through many hands. Many of the problems of data resolution, surrogacy and timeliness are today less problematic, yet more data are second hand and more data are collected using unscientific research designs (indeed they are often not principally collected for 'research' at all).

## 2. The developing digital data infrastructure

### 2.1 Changes in supply, pricing and access

In physical and social science alike, the costs of data have generally been a (sometimes the) major component of the costs of GIS creation. The order of magnitude of data costs reflects a number of technological and secular imperatives which govern the supply, pricing, and access aspects to data availability.

In the early days of GIS, the data bottleneck of (manual or semi-automated) digitising presented a major impediment to the creation of spatially-referenced databases, particularly if the hard copy source documents were complicated or ambiguous. Early software systems provided (by present day standards) fairly unsophisticated procedures for detecting and correcting the results of error-prone digitising. Moreover, framework spatial data, such as those created and maintained by national mapping agencies were available only in hard-copy printed form, and in the early days of GIS there was resistance to initiating the task of converting legacy hard copy maps to digital form.

A wealth of digital data has since come into existence. First, and as with computer hardware, new technology is playing an important role. In particular, the wide (selective) availability of global positioning systems makes creation of new digital datasets much more straightforward than hitherto. Second, most national mapping agencies have gradually overcome their initial reluctance to create digital versions of their paper records, while at smaller scales private providers have created a range of digital atlas products. And third, computerised logging of the physical and social environment takes place with ever-increasing frequency, and to ever-greater levels of detail: for example through high-resolution remote sensing of the physical and built environments and the digital encoding of consumer purchasing behaviour (through loyalty programmes and the development of relationship marketing) in the socioeconomic realm.

Yet this has not created a panacea for data modelling. In practice, accurate field recording of data remains an expert task and sound geographical analysis presumes sound data standards. Many national mapping agencies (such as Great Britain's Ordnance Survey) have only succeeded in going digital in the face of increasingly stringent public expenditure constraints by recovering vastly increased proportions of their creation and maintenance costs through user charges: the inevitable consequence is a rationing of framework data on an ability to pay basis. Similarly hawkish data pricing regimes may apply to the data products from the new generation of high-resolution satellite sensors, while high royalty charges dissuade many business users from census data and census data products in some countries (such as the UK). At the same time, governments are reluctant to fund even their traditional linear project design-driven surveys, in view of the apparent tide of information created using new data capture technologies. With respect to the academic realm, the rise of interdisciplinary science is leading to a higher incidence of jointly-funded projects, and the commonplace situation in which the creators of spatial data may be widely separated from some of the communities of end users. As creators and users of data become more and more separated, in space, time, and intellectual tradition, the ability to describe data becomes increasingly critical. The creator must be able to tell the user about methods, accuracies, formats, and all of the details needed to transfer, open, and make effective use of the data. Moreover the user must be able to determine whether a given data sets meets or falls short of requirements, and this is increasingly accomplished through metadata.

### 2.2 The changing remit and requirements of modelling

The early years of the spatial analysis paradigm were associated with the development of wide-ranging models of physical and social systems. The remit of such models was avowedly ambitious, yet on reflection the data infrastructure was not commensurate with the tasks in hand. A number of commentators have identified reasons for the subsequent demise of large scale socioeconomic modelling activity, although the innovation of GIS has brought with it a renaissance in model-building activity. Moreover, any decline in large-scale modelling of socio-economic systems has been matched by the rapid growth of environmental modelling, much of it coupled with or otherwise making use of GIS.

The new is quite different from the old, however. Within the socioeconomic realm, Birkin (1996) has described how the current generation of spatial interaction models, for example, seeks only to model limited (in terms of spatial extent, time frame and attribute range) aspects of urban sub-systems. This in part reflects secular trends in all developed societies away from system-wide planning, yet it also reflects a profound reappraisal of what we now consider to be the appropriate domain and capability of analytical models. Today's urban models are much more data-rich in two respects. First, the revolution in the supply and availability of geographical information means that data no longer represent coarse zonal aggregations, and thus that the data model of spatial distributions bears a closer correspondence with reality. Second, the first generation of urban models used data derived exclusively from public sector sources and which were thus restricted to the limited range of variables of interest to officialdom. Whilst such data can be used, singly or in combination, to create crude indicators of human behaviour and activity patterns, such indicators bear at best a very imperfect correspondence with reality.

Within the socioeconomic realm, the present status of modelling is rather ambiguous. Within academia, disenchantment with urban modelling leaves it as an area of activity with a significantly reduced real share of intellectual activity compared to, say, twenty years ago. Business applications of data-rich partial models of components of urban systems are buoyant, and today client repeat purchases provide vindication of the validity of spatial interaction and other modelling approaches. Within planning, there has never been a greater need for accurate data and analytical models of urban systems, because the rate, scale, and pace of change has never been greater. Yet, in the UK at least, there is disquiet about the "predict and provide" approach to planning which has hitherto been based upon aggregate modelling approaches.

### 2.3 Model linkage: towards a new perspective?

The linear project design presumed that resources were available for a linear, vertically integrated sequence of events. Today's research environment is much less straightforward. The strictures of public expenditure make it less likely that large-scale purpose-specific research will be funded, while information commerce makes it less than unequivocal that the best secondary data will be available. Yet data warehouses are bursting with data that might be combined to create richer profiles of landscapes, morphologies, households, and activity patterns than have ever been created before. While the developing geocomputation paradigm presents us with some brute force mechanisms for searching out generalisations from large and complex datasets, we may have no way of knowing whether such generalisations hold any scientific validity.

A negative view of this research environment would suggest that a price has been put on scientific truth that lies beyond the budget of many researchers. There is some truth in this, yet economic imperatives need also to be viewed in their technological context. In truth, as our retrospective of urban modelling above has illustrated, data collected through the linear project design did not provide a panacea in practice. Today's digital data infrastructure is more detailed, relevant, and up-to-date than ever before. The problem is that this infrastructure is also more piecemeal, and hence possibly ill-founded and unsafe.

The environment for spatial analysis is GIS, which has always been an applications-led technology. The sophistication of current applications requires a breadth and depth of data that could never have been sustained by established data collection methods. Today's open and desk-top GIS alike are geared towards the analysis of application-specific “”horses for courses” datasets. Such datasets are required to model real-world systems that are dynamic and fast-changing, and thus the timescale between data collection and availability of secondary analysis needs also to be shortened. Our understanding of physical and social systems alike is now of such sophistication that infrequently collected, aggregate, and surrogate spatial data are simply not good enough. These are all crucial considerations, yet they all lie outside the remit of the linear project design. Are we therefore faced with a stark choice between scientific validity and making do with inappropriate, overly-aggregate, out-of-date indicators? The rejection of Census-based geodemographics in favour of lifestyles (i.e. data warehouse) analysis in much of business geographics suggests that the road to scientific truth is no simple one-way street, and that proponents of inductive data-led thinking have their supporters in the world of application.

Framed in these terms, one of the big questions for GIS at the turn of the millennium must be: Can the new digital data infrastructure be assembled together in a sufficiently accurate, orderly and rational way to bridge relevance, richness and academic respectability? Goodchild and Longley (1999) use the term “concatenation” to describe the integration of two or more different data sources, such that the contents of each are accessible in the product. The polygon overlay operation is one simple form of concatenation. They use the complementary term “conflation” to describe the range of functions that attempt to overcome differences between data sets, or to merge their contents. Conflation thus attempts to replace two or more versions of the same information with a single version that reflects the pooling, or weighted averaging, of the sources.

## 3. Model linkage in practice

### 3.1 RS GIS concatenation

Census information and satellite imagery are diverse sources of information. Longley and Mesev (1997) use information from the 1991 UK small area census statistics as ancillary information to improve the classification accuracy of a contemporary (LANDSAT TM) image of Bristol. Information from the Census is used to assist in sample training and post-classification sorting. The resultant hybridised dataset is designed with a specialised purpose in mind: to provide detailed data models of the distribution of population and domestic property. This is used to reappraise conventional analysis of the density at which urban space is occupied and through comparisons Longley and Mesev (1997) develop density gradient profiles for different categories of urban space filling, such as built form, residential, households, and population. They demonstrate that the differences between these apparently similar categories are more than semantic, and can heavily condition whether and to what extent we might consider density profiles characteristic of particular settlement types. The optimistic message of this work is that, once the differences between different conceptions of urbanity have been clearly grasped, it is possible to develop a range of customised indicators of urban morphology. In this way, customised GIS-based data models are informing our thinking about the ways in which urban settlements fill space, as well as providing detailed information as to the morphology of particular settlement structures.

### 3.2 Conflating geodemographics and lifestyles

Lifestyles is a broad term that has been used to describe data pertaining to the consumption of a wide range of goods and services by identifiable individuals and households. Lifestyles data originate from a diverse range of sources, such as guarantee card returns, questionnaires attached to nationally circulated prize draw entries, and market research surveys. They are usually georeferenced through the postcode system (e.g. in the UK to the unit postcode, which typically comprises 15 or so addresses in urban areas). At least one UK data warehouse estimates that it holds up-to-date information on 11 million UK households. Such data have evident use for direct marketing, for past consumption habits are key guides to future behaviour. Harris (1999) has analysed the anonymised individual/household records from one particular lifestyles questionnaire which was mailed out in October 1996. The number of respondents to this survey constitutes 10.8% of all households in Bristol, UK (population 636,000): this makes the survey larger in size than a mini census, yet the characteristics of non-respondents are likely to be very unrepresentative of respondents. In recent years, lifestyles approaches have gained some ground as tools for geomarketing at the expense of the use of census and composite geodemographic indicators, because the latter are increasingly out of date (the last UK Census was held in 1991), they are expensive to use because of UK royalty structures and, perhaps most damning of all, the census contains too few variables that bear an identifiable correspondence with consumer behaviour (most notably in the UK, because of the absence of an income question in the Census).

The geodemographics lifestyles debate thus epitomises the tensions described in Section 2 above. Geodemographics is based on tried and trusted techniques and derives from a dataset (the Census) which has been designed and implemented using the most rigorous research design principles; and yet at the end of the day, it is out of date, and can supply at best only very imperfect indicators of real-world consumer behaviour. Sampling theory tells us that reweighting of largely self-selecting samples on the basis of sub-group response rates is foolhardy; yet survey research practice tells us that quantitative indicators should be direct and transparent, and that survey results are only directly applicable to the population from which the respondents were drawn (few of us would wholly identify with our digital past-selves who filled out a census form at the start of this decade).

A middle path between these two lies in Batey and Brown's (1995) assertion that lifestyle descriptors can be used as a wrapper to add depth to the labels assigned to different geodemographic groups. Thus, for example, the SuperProfiles category "affluent achievers" has fairly distinctive Census characteristics in terms of house construction type, socio-economic status and car ownership, to which lifestyle labels about theatre and restaurant patronage, share registers, newspaper readership, and credit card usage are added. The data from which these labels are obtained are in many cases collected by unscientific means or strictly pertain only to coarser aggregations of households. Yet Harris's (1999) cluster analysis of (unweighted) lifestyle data finds some practical validity to this approach: it nevertheless runs rough-shod over conventional views about how scale and aggregation issues should be tackled.

## 4. The future of spatial analysis

Goodchild and Longley (1999) suggest that the kinds of circumstances and imperatives presented in the preceding discussion will lead to the emergence of the following kinds of spatial analysis in the coming years:

- analysis of data whose meanings are clearly understood, making it easier for multidisciplinary teams to collaborate;
- analysis of data which are routinely collected in the day-to-day functioning of society and the everyday interactions between humans and computers;
- analysis of data with widespread use, generating demands that can justify the costs of creation and maintenance;
- analysis of data with commercial as well as scientific and problem-solving value, allowing costs to be shared across many sectors;
- methods of analysis with commercial applications, making it more likely that such methods will be implemented in widely available form;
- methods implemented using general standards, allowing them to be linked to other methods using common standards and protocols.

This statement has highlighted the way in which the advanced information economy of the late 1990s has multiplied the number of potential sources of (rich) digital information, yet in ways which will be less standardised and project-specific than those implied by the linear project design. A major challenge to the GIS community is to devise methods to reconcile diverse datasets with different data structures or spatial referencing systems. Only in this way will GIS be able to tease out the complex relationships that exist between projects, data sets, and analytic techniques in modern science. The self-perception of rigour amongst spatial analysts has hitherto been misplaced because of the vagaries and inadequacies of data quality, resolution and richness: progress requires us to face up to the fact that the linear project design was never a panacea in practice.

Batey P, Brown P (1995). From human ecology to customer targeting: the evolution of geodemographics. In Longley P, Clarke G (eds) **GIS for business and service planning.** Cambridge, GeoInformation International: 77-103

Birkin M (1996). Retail location modelling in GIS. In Longley P A, Batty M (eds) **Spatial analysis: modelling in a GIS environment.** Cambridge, GeoInformation International: 207-25

Curry M R (1995). GIS and the inevitability of ethical inconsistency. In Pickles J (ed.) **Ground truth: the social implications of geographic information systems**. New York: Guilford Press: 68-87

Goodchild M F, Longley P A (1999). The future of GIS and spatial analysis. In Longley P A, Goodchild M F, Maguire D J, Rhind D W (eds) **Geographical information systems: principles, techniques, management and applications**. New York, Wiley: 1: 567-80

Harris R (1999). **A comparative analysis of a lifestyle and geodemographic typology.** Working paper. Bristol, University of Bristol

Kennedy P (1979). **A guide to econometrics**. Oxford, Martin Robertson

Longley P A, Mesev V (1997). Beyond analogue models: space filling and density measures of an urban settlement. **Papers in Regional Science** 76: 409-27

Martin D J (1996). **Geographic information systems: socioeconomic applications**. London, Routledge

Valavanis S (1959). **Econometrics.** New York, McGraw-Hill

**Duane F. Marble**
Center for Mapping & the Department of Geography, The Ohio State University,
Columbus

## The Future of Spatial Analysis and Geographic Information Systems

It is certainly time, if not past time, to examine in some depth both existing and potential relationships between that portion of geography that we call spatial analysis and the complex technology that has become labeled geographic information systems. It is clear in my own mind that future developments in both areas are going to require higher levels of integration than we see now if both are to advance significantly.

One problem that I have encountered recently seems to arise out of some current operational definitions of spatial analysis. Some forms of what I would personally identify as spatial analysis are playing a major role in GIS applications today (e.g., transportation optimization models). Strangely enough, recent conversations I have had with some prominent geographers who identify themselves as spatial analysts reveal that they consider these activities to be modeling and not really spatial analysis. I conclude from these informal conversations that “real” spatial analysis is somehow viewed as deeply rooted in spatial statistics and that non-statistical, analytic methods are considered to be outside of spatial analysis. My personal view is that both spatial statistics and analytic spatial models (both descriptive and optimizing forms) are simply complementary aspects of what should be a more comprehensive working, and generally accepted, definition of spatial analysis.

If we accept this broader definition of spatial analysis, then the question becomes why are some aspects of it widely accepted in the GIS area while others are not? The answer does not lie in a lack of participation by geographers or a lack of knowledge of spatial analysis by GIS developers as Longley and Batty (and others) have suggested.[2] Rather, it may be found in the understandable commercial orientation of the companies developing GIS technology. The incorporation of spatial optimization models into the GIS has resulted in massive cost savings by users of the technology that has, in turn, led to demand for more sophisticated and easier to use tools in this area. It has yet to be clearly shown that the incorporation of other forms of spatial analysis would generate a similar level of utility to institutions regularly dealing with complex, real-world spatial problems.

It is my belief that other forms of spatial analysis could make similar contributions if (a) an effective attempt were made to establish their clear utility within the context of large-scale, real-world problems, and (b) their present, traditional view of spatial problems is replaced with a broader and more realistic one. Let me comment on each of these in turn. My first comment relates to the creation of a demand among the rapidly increasing GIS community for the results that can be provided by the tools in question (e.g., spatial statistics). The problem here is not unlike the map projection problem that has dismayed the GIS community for years. Basically the user’s question has been:Why bother with map projections? Things work OK if I ignore them. Even self-styled GIS consultants were telling clients how much money they could save in data conversion of they just forgot about all that stuff. It has now been generally demonstrated that such an approach leads to expensive errors. We need to demonstrate that the other components of spatial analysis (as contrasted to the highly successful spatial optimization models) can make a real difference. It will not be easy, and is unlikely to lead to many academic brownie points, but it can be done.

The second point is one that I have brought up on several previous occasions. Because of the difficulties involved in both conceptual structures and computation, we have elected to deal with representations of the world around us that are far too limited. For example, many of our spatial views are one-dimensional in nature where the space equals distance assumption is so ingrained that it is never even mentioned. What about the other dimension of the two-dimensional space that we contend we deal with? Direction is generally ignored in geography but if we look around, we see that directional statistics play an important role in other disciplines such as ecology, oceanography, and geology.[3] We have also generally neglected the role of time in our work in spatial analysis. In reality, what are called for are powerful spatial models and tools that work in a multi-dimensional space-time framework. It is my contention that melding existing spatial analysis approaches with GIS represents the single most important action that will permit significant developments to take place in this direction.

Operationally, I believe that the spatial analysis community urgently needs to follow the lead of the GIS community into an object-oriented view of how they carry out their work. Four years ago a colleague of mine, Dr. Randy Jackson, presented a detailed and cogent argument for the utility of object-orientation in Regional Science. He made this argument prior to the recent, major move of GIS developers into object-oriented development tools (e.g., ESRI's Map Objects and Net Engine). This move has created a major change in GIS technology and it is clearly time for the spatial analysis community to think very seriously about Jackson's seminal proposals.[4]

In conclusion, I feel that both spatial analysis and geographic information systems are on the brink of a major revolution that will lead to substantial increases in the scope and power of both areas. This will not be successful unless the two areas become much more highly integrated both conceptually and operationally. I would hope that the forthcoming meeting would lead to such an integration.

1. A brief position paper prepared in conjunction with the forthcoming 1998 Varenius Workshop on Status and Trends in Spatial Analysis.

2. Paul Longley and Michael Batty, 1996. *Spatial Analysis: Modelling in a GIS Environment*. Cambridge: GeoInformation International.

3. Marida, K.V., 1972. *Statistics of Directional Data*. New York: Academic Press is a classic reference and Gaile, Gary C., 1980,*Directional Statistics*. Norwich: GeoAbstracts summarized this area for geographers. Little interest has developed, perhaps due to a lack of relevant theory in human geography.

4. Jackson, Randall W., 1994. Object-oriented modeling in Regional Science: an advocacy view, *Annals of Regional Science*.

**Harvey J. Miller**
Department of Geography, University of Utah, Salt Lake City

## Beyond the Isotropic Plane: Towards a Geospatial Analysis

### Introduction

A gap exists between the conceptualizations of geography in spatial analysis and the models of geography available in geographic information science (GISci). To a large degree, spatial analysis greatly simplifies its central object of study, i.e., geographic space. In the past, this simplification was forgivable due to the limitations of traditional data handling methods. However, the rapidly improving ability to handle complex and voluminous geographic data requires a re-examination of spatial analysis from the "ground up." A geospatial analysis would examine convenient spatial assumptions and, where appropriate, expand these representations using the computable models of space, spatial objects and spatial relationships available in GISci. This will result in a geographically-sensitive spatial analysis that could revolutionize theories of spatial processes and enhance the relevance of spatial analytical techniques to real-world decision making.

### Spatial Analysis

An appropriate point-of-departure is to define "spatial analysis." Although several definitions are available, Nyerges (1991) defines a set of "analytical map questions" that highlight the role of geographic space, geographic objects and geographic relationships. Questions dealing with location and extent include "Where is the phenomenon of interest?" and "Why is the phenomenon there?" Questions dealing with distribution and pattern concern the existence and nature of regularity in the phenomenon's spatial distribution. Spatial association questions concern coincidence among two or more phenomena in space. Questions dealing with spatial interaction focus on the existence and nature of linkages among spatial phenomena. Finally, spatial change questions address the geographical change in a phenomenon and the factors that cause that change.

With these analytical map questions in mind, consider the standard representation used in spatial analysis. With some exceptions (to be noted later), the "standard model" consists of an isotropic plane containing points representing geographic objects. Lines, areas and surfaces are sometimes used in spatial statistics (e.g., Getis and Boots 1978). However, it is rare for these objects to be "mixed" so that each geographic object class is appropriately represented; instead, a single representation is usually imposed on all geographic objects. Relationships among these objects typically reduce to a single dimension, namely, distance or some function of distance (time, cost).

Since geographic representations condition the answers we can obtain from analytical map questions, a critical research question is the extent to which geographic representations in spatial analysis have affected our understanding of spatial phenomena and our prescriptive solutions to geographic problems.

## Towards a Geospatial Analysis

The rapidly improving capabilities of geographic information technologies along with the increasing availability of detailed digital geographic data can greatly reduce the "impedance mismatch" between geographic reality and formal geographic representations in spatial analysis. A geospatial analysis will examine the convenient geographic assumptions used in spatial analysis and, where appropriate, replace these representations with computable models of geography available in GISci.

Some precedents for a geospatial analysis exist in the spatial analysis literature, particularly within the fuzzy boundary between spatial analysis and GISci. These include the following:

Geographic Space. Terrain can substantially influence the locations of facilities and transportation routes as well as human and physical interactions across space. Goodchild (1977), Goodchild and Lee (1989) and Lombard and Church (1993) develop facility location and routing models that exploit terrain information in digital elevation models. Land use/land cover can have similar effects. Werner (1968) formulates a computational procedure for capturing the "refracting" effect of cost polygons that reflect physical characteristics on transportation routes. Golledge et al. (1969) and Tobler (1976) develop formal representations of perceived/experienced space; these can greatly enhance the analysis of human spatial behavior (see also Egenhofer and Mark 1995). Spatial analysis at regional, national and international scales requires an explicit treatment of the Earth as a sphere. Love, Morris and Wesolowsky (1988) discuss facility location on a sphere while Raskin (1994) discusses more general spatial analytical techniques (including interpolation) on a sphere. Time is also tightly interlinked with geographical space and individuals' perceptions and actions (Egenhofer and Mark 1995; Hägerstrand 1970). Some recent efforts have been directed at implementing space-time frameworks within spatial analysis (Kwan 1998; Miller, 1991, 1998)

Geographic Objects. Miller (1996) discusses expanding location models to include other spatial objects besides points to represent facilities and clients. Okabe and Saddahiro (1994) and Okabe et al (1995) develop spatial statistical procedures that relate points to surfaces and points to networks, respectively. Okabe and Miller (1996) provide support for mixing spatial objects in spatial analysis by developing computational procedures for measuring average, minimum and maximum distances among all possible pairings of points, lines and polygons. Several researchers have developed shape measures (e.g., Boyce and Clark 1964; Massam and Goodchild 1971; Moellering and Rayner 1981; Tobler 1978) for describing and comparing complex geographic objects. Fractals have been used to describe morphology and processes in both physical and human geographic phenomena (see MacLennan et al. 1991).

Geographic Relationships. Other factors besides distance can affect geographic relationships, either independently or in conjunction. For example, both distance and direction can influence knowledge of an environment (e.g., Moore 1970). Peuqeut and Xiang (1987) develop a procedure for computing directional relationships between planar-embedded polygons. Humans often use toplogical relationships among spatial objects as their primary information about geography, with distance and shape information being secondary (Egenhofer and Mark 1995). Formalisms for capturing topological relationships have been developed (e.g., Egenhofer and Franzosa 1995; Egenhofer, Mark and Herring 1994). Boundaries (physical, built, administrative also condition geographic relationships, e.g., the geometry of boundaries can strongly influence spatial interaction models (Griffith 1982). The geographic scale at which behavior occurs can condition all of the relationships mentioned above, at least with respect to human behavior (Gale and Golledge 1982).

## A Research Agenda

The very brief review above indicates that the elements of a geospatial analysis exist in the spatial analysis and GISci literatures. However, these efforts are piecemeal, isolated and not well integrated into mainstream spatial analysis. Required in a comprehensive and rigorous framework to integrate existing and potential geospatial analyses.

The ordering principle for examining spatial analytical problems should be "What is the best way to represent geography?" based on the empirical nature of the geographic space, geographic objects and geographic relationships in the problem domain. After classifying problem domains based on best representations, the current "state-of-the-practice" within each domain should be assessed, i.e., how is geography represented in practice? Systematic differences between the best and current representations must be assessed carefully. Note that this approach treats the best representations as the "gold standard" and views current practice as deviations from that standard.

After determining best representations and typical deviations used in practice, the next questions concern the costs of these deviations and the benefits of achieving more realistic geographies within each problem domain. Is it worth improving geographic representation in the particular problem domain? What is achieved with respect to theoretical insights and improved prescriptive modeling? These questions must be answered in light of the state-of-the-art in GISci and GIS software.

**Literature Cited**

Boyce, R. and Clark, W. (1964) "The concept of shape in geography," Geographical Review, 54, 561-572.

Egenhofer, M. J. and Franzosa, R. D. (1995) "On the equivalence of topological relations," International Journal of Geographical Information Systems, 9, 133-152.

Egenhofer, M. J. and Mark, D. M. (1995) Naïve Geography. National Center for Geographic Information and Analysis, Technical Report 95-8.

Egenhofer, M. J., Mark, D. M. and Herring, J. (1994) The 9-Intersection: Formalism and its Use tor Natural-Language Spatial Predicates. National Center for Geographic Information and Analysis, Technical Report 94-1.

Gale, N. and Golledge, R. G. (1982) "On the subjective partitioning of space," Annals of the Association of American Geographers, 72, 60-67.

Getis, A. and Boots, B. (1978) Models of Spatial Processes: An Approach to the Study of Point, Line and Area Patterns, Cambridge: Cambridge University Press.

Golledge, R. G., Briggs, R. and Demko D. (1969) "The configurations of distances in intra-urban space," Annals of the Association of American Geographers, 85, 134-158.

Goodchild, M. F. (1977) "An evaluation of lattice solutions to the problem of corridor location," Environment and Planning A, 9, 727-738.

Goodchild, M. F. and Lee, J. (1989) "Coverage problems and visibility regions on topographic surfaces," Annals of Operations Research, 18, 175-186.

Griffith, D. A. (1982) "Geometry and spatial interaction," Annals of the Association of American Geographers, 72, 332-346.

Kwan, M. (1998) "Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework," Geographical Analysis, 30, 191-216.

Lombard, K. and Church, R. L. (1993) "The gateway shortest path problem: Generating alternative routes for a corridor location problem," Geographical Systems, 1, 25-45.

MacLennan, M, Fotheringham, A. S., Batty, M. and Longley, P. A. Fractal Geometry and Spatial Phenomena. National Center for Geographic Information and Analysis Technical Report 91-1.

Massam, B. H. and Goodchild, M. F. (1971) "Temporal trends in the spatial organization of a service agency," Canadian Geographer, 15, 193-206.

Miller, H. J. (1991) "Modeling accessibility using space-time prism concepts within geographical information systems", International Journal of Geographical Information Systems, 5, 287-301

Miller, H. J. (1996) "GIS and geometric representation in facility location problems," International Journal of Geographical Information Systems, 10, 791-816.

Miller, H. J. (1998) "Measuring space-time accessibility benefits within transportation networks: Basic theory and computational methods," Geographical Analysis, in press.

Moellering, H. and Rayner, J. N. (1981) "The harmonic analysis of spatial shapes using dual axis Fourier shape analysis (DAFSA)," Geographical Analysis, 13, 64-77.

Moore, E. G. (1970) "Some spatial properties of urban contact fields," Geographical Analysis, 2, 376-386.

Nyerges, T. L. (1991) "Analytical map use," Cartography and Geographic Information Systems, 18, 11-22.

Okabe, A. and Miller, H.J. (1996) "Exact computational methods for calculating distances between objects in a cartographic database," Cartography and Geographic Information Systems, 23, 180-195.

Okabe, A. and Sadahiro, Y. (1994) "A statistical method for analyzing the spatial relationship between the distribution of activity points and the distribution of activity," Geographical Analysis, 26 152-167.

Okabe, A., Yomono, H. and Kitamura, M. (1995) "Statistical analysis of the distribution of points on a network," Geographical Analysis, 27, 152-175.

Peuquet, D. J. and Xiang, Z. (1987) "An algorithm to determine the directional relationship between arbitrarily-shaped polygons in the plane," Pattern Recognition, 20, 65-74.

Raskin, R. G. (1994) Spatial Analysis on the Sphere: A Review, National Center for Geographic Information and Analysis, Technical Report 94-7.

Tobler, W. R. (1976) "The geometry of mental maps," in R. G. Golledge and G. Ruhston (eds.) Spatial Choice and Spatial Behavior, Columbus, OH: Ohio State University Press, 69-82.

Tobler, W. R. (1978) "Comparison of plane forms," Geographical Analysis, 10, 154-162.

Werner, C. (1968) "The law of refraction in transportation geography: Its multivariate extension," Canadian Geographer, 7, 28-40.

**Morton E. O'Kelly**
Geography, Ohio State University

*Preface: what follows is a set of comments, notes, and examples, relating to some of the broad questions in the call for proposals. These are intended as personal opinions, and discussion points, and differ somewhat in format from some of the other position papers.*

## Has GIS been successful at making spatial analysis widely available to physical and social scientists?

Many commentators have noted the mismatch between the sophisticated capabilities of Spatial Analysis/GIS and the techniques that are actually employed in practice. This mismatch is especially apparent in the case of social science applications. Thus, while physical scientists have quite naturally sought out or developed new tools (DEM, spatial prediction via kriging, landform modeling and so on) my perception is that social scientists have not been similarly empowered by GIS. This contrast is not surprising, of course, given the geo-science basis for many of the techniques in spatial statistics. Nevertheless, advanced spatial modeling, visualization and generalization are typically not used as much as they could be in desktop demographic mapping applications. Too often business geographic presentations extend no further than address matching, point-in-polygon operations, and choropleth maps with basic demographic variables. However, technical capabilities are expanding to include network analysis, spatial interaction models and optimization, for example. GIS provides a superb platform for integrating spatial data, and with appropriate techniques these data may be used to advantage in social science modeling. It should be apparent, however, that geographical and statistical analysts need to do some work to develop and transfer new techniques to the real-world (e.g. for spatial prediction). This paper fleshes out this argument with reference to some of the themes suggested as discussion points for this workshop.

Increasingly, large volumes of disaggregated individual level data are available to the analyst [see e.g. Kwan's statement for this workshop]. Then, by employing a moderate amount of aggregation, it is possible to derive a spatially referenced data base with a common spatial context. This would seem to provide a platform for techniques such as multi-level models (Kelvyn Jones et al; and Cressie's statement for this workshop). I would like to explore this potential question at the workshop. In my own experience, for example, in retail trade area analysis, I have been combining observations into blocks to provide a number of observations of individuals with more or less common retail choice sets. These commonalties may be exploited to great effect. In situations where a residential zone is accessible to several alternative destinations, we may use the variations in rates of patronage to the several alternatives to estimate the impact of size, distance, spatial structure effects, and other commonly used explanatory variables that are typical of the spatial interaction modeler's arsenal. The paper discusses techniques such as density models, interaction models, and so on, and outlines the appropriate estimation steps needed to fit parameters in these models. While the comments are made in the context of a specific practical application, the relevance of these techniques to other problems (hospital planning, participation in social programs, and school assessment) is fairly obvious.

**An idea for improving the environment currently provided by GIS**

GIS/spatial analysis projects focus a lot of attention on discussions of graphical user interfaces (GUI), menu layout, and ease-of-use issues. The discussion can drift into the appearance of dialog boxes, the choices of selection sets, and the offering of a variety of alternative objectives and constraint formulations to the clients. This focus would make more sense if the underlying data structures and models and algorithms were already fully understood and worked out, but regrettably, the basic methodological issues are still in need of intensive effort. More important that these usability issues would be expert system support from a knowledge base that embodies experience, best practice, and even rules-of-thumb. For instance, when people discover spatial analysis via a GIS package, they encounter a very steep learning curve; (e.g. gravity interaction models in ArcInfo, traffic assignment models in Transcad, or Kernel density estimation in ArcView Spatial Analyst). My suggestion is that the software environment should provide help and give substantive guidance to non-specialists (and “learners”). This, in my view, would be a major improvement over the current state of knowledge. Ideally, software for GIS/spatial analysis would be used by people with a thorough exposure to, and training in, geographical analysis. In reality though, spatial analysis concepts may be completely unfamiliar to those who have access to GIS. To give a few short examples that would be worth fleshing out further at the workshop:

- A menu for kriging may put powerful tools at the disposal of a user: if that user does not appreciate some of the required properties of the theoretical covariogram, nonsense can result. The situation could be improved by giving the user some support in terms of fundamentals.
- As an over-simplified example (just for the sake of illustration): in trip distribution models, a user intending to use a value of a parameter equal to 0.4, would be informed that this value implies an average journey to work length of 35 miles. This is the kind of consistency check, and pre-estimation, and ideally verification and validation that we expect people to do with more routine statistical analysis and should be used as more complex techniques become available.
- Another example would be a warning of the need for edge correction to a user about to estimate an empirical K(h) function, where h could be up to 50 km, in a study region of say 100 square km.

**Diversion of effort away from fundamental research**

The goal differences between the research community and the corporations and individuals who design software for applications purposes are fairly obvious. Thus, in my opinion, there is a tension between GIS design and creative mathematical/spatial analysis. The GIS design process, has as its goal the efficient and effective application of existing technology to the problem set (Marble). For all its merits, and for all its success in preventing horror stories when implemented rigorously, it is clear that GIS design addresses a question that is much different from the creative process of new model development. The design protocol/regimen requires that the analyst make successively more specific passes at the specification of a solution to the problem. Knowing who is going to use the system, and what the system is to be used for, is rightly given priority in such a scheme. Research*per se*, and extension of the state-of-the-art is not the goal, although research extensions could occur as by-products from a particular application. The demand for skilled individuals to do this kind of work for software companies will mean the reduction of the pool of people ready and willing to do much-needed fundamental academic research. The competition for scarce talent in this area has already been felt in the job market. There are exceptions, of course, and many of those who have successfully straddled both sides of this fence will be in attendance at the workshop, and so I hope to hear more examples and feedback on this discussion point.

## Some thoughts on how we proceed from here

It is probably worth exploring the changing labor/capital intensity of inputs to GIS and spatial analysis research. In the 60s, quantitative spatial analysis was a time consuming labor intensive activity, with the resultant product regarded as a research work because of the time and effort needed to make it. Nowadays routinization has made many analytical steps much easier, and we could realistically expect a powerful data base manager, a good statistical analysis package, a GIS mapper, and perhaps a sophisticated report writer to produce custom reports for 100 MSAs in the USA. Although some technical skill would be needed to do the data integration steps, the products of this process would not be generally acceptable as research.

The archtypical example is the suite of tools for demographic data mapping. These data CD ROMs come packed with data for hundreds of undigested variables and allow the user to select infinitely varied study areas. Products such as Census-CD, a simple desktop thematic mapper, is capable of producing an immense array of maps and we have to ask if we have taken a step backwards in making the production of these maps so easy: we give people/end users access to reams of undigested data and expect them to be able to make intelligent use of these covarying data sets. Didn’t the factorial ecologists teach us to boil the data into essential factors?

The correct model, to my mind, is one of continuous improvement. An operational version of an idea should be rapidly prototyped, using either novel or existing text book methods. The tool is presented, tested, and debugged. Then a series of upgrades, re-writes, enhancements, and so on are built. These are upgrades both to the way that the simple model is implemented, but also perhaps, new discoveries of critical process and adaptations. An example that typifies the successful idea here might be the PASS dial-aride software system, one which has many geographical ingredients, complex data base linkages, and a challenging underlying algorithmic problem (vehicle routing with time windows). Another example might be the continuous improvement and re-refinement of the location-allocation suite of models, which some here will remember fondly from the mainframe days at IOWA and the ALLOC package. A final example might be in the area of trade area mapping and estimation of gravity model parameters.

Of course the relevance of these ideas must depend somewhat on the position one holds in the spectrum of pure research through to applied commercial software development. I'm coming at this from the point of view of someone who is quite comfortable experimenting with ideas and in thinking about general new ideas for spatial analysis. Often such ideas are exploratory, or are left partially documented, perhaps to be revisited at a later time. I would find it difficult to change hats and consider application issues, because I prefer to think of spatial analysis as an intermediate calculation on the way to exploration of actual processes.  This academic/practitioner division of labor that has served us well so far: a question for the workshop is whether the future growth of spatial analysis and GIS needs a revised model.

**Atsuyuki Okabe**
Center for Spatial Information Science, University of Tokyo, Tokyo

## Toward a Toolbox for Analyzing Spatial Relations

Since the early 90's, a general framework for analyzing spatial relations has been intensively developed by Egenhofer and others. His model, called the 9-intesection model, is very useful for analyzing topological relations between objects, but the method is a little bit far from the analysis of spatial relations observed in geographical phenomena, such as the analysis of spatial relationship between the distribution of pine trees and that of cider trees.

The objective of our study is, first, to provide a general framework for analyzing spatial relations observed in geographical phenomena, and second to develop a user-friendly toolbox of methods classified under this general framework. The implementation of this toolbox will be achieved in a GIS environment.

The toolbox consists of small boxes, which are labeled according to geometrical shapes of objects. We classify geographical objects by points, lines and areas (polygons), and in terms of these objects, we classify spatial relations into 6 relations: points-points, points-lines, points-polygons, lines-lines, lines-polygons, polygons-polygons.

1. Points-Points. Examples are: a spatial relationship between the distribution of railway stations and that of factories; that between distributions of two kinds of species (in ecology). A fairly many methods have been developed in geography and ecology, which are called association methods.

Strangely enough, methods for analyzing other spatial relations have been less developed in the related literature.

2. Points-Lines. An examples is a spatial relationship between the distribution of withered tress with respect to a network of express ways.

3. Points-Polygons. An example is a spatial relationship between the distribution of high-rise apartment buildings around big parks.

4. Lines-Lines. An example is a spatial relationship between a network of canals and that of roads.

5. Lines-Polygons. An example is a spatial relationship between the distribution of rice fields with respect to rivers.

6. Polygons-Polygons. An example is a spatial relationship between the distribution of race fields and that of wheat fields.

Methods for spatial relations, in particular, 2-6, should be developed. One possible and promising method would be the use of generalized Voronoi diagram, or more specifically, Voronoi diagrams for points, lines and polygons. A few initial attempts are shown in Okabe, Boots and Sugihara (1992).

**Pat A. Pellegrini**
Department of Geography, The Ohio State University, Columbus

From the perspective of a scientist involved in spatial analysis, it is clear that GIS has been reasonably successful at making spatial analysis more widely available to physical and social scientists. More specific to my own research in spatial analysis, the field of spatial point pattern analysis has witnessed substantial advances over the last twenty years (Bailey and Gatrell 1995; Cressie 1991; Diggle 1983). To be sure, the recent interest in spatial point pattern analysis has been fueled in part by the interest spawned more generally in spatial statistical methods as a result of the development and use of GIS for the storage, manipulation, visualization and analysis of spatially referenced data.

However, as noted by several practitioners (Goodchild 1987; Openshaw 1991), commercial GIS products typically provide little in the way of statistical functionality and so the computational implementation of spatial point pattern analysis methods has relied on custom packages such as INFOMAP (Bailey and Gatrell 1995) or on "libraries" of spatial statistical routines which are accessed through statistical packages (eg., SPLANCS and S-plus) or through GIS packages (eg., Spatial Analysis Toolkit and ARC/INFO). Clearly, the environment currently provided by GIS is deficient and must be augmented with user-written code.

To improve existing GIS functionality with regards to spatial point pattern analysis, and more generally, spatial analysis, we need to consider the methodological insights that may gained by embedding the appropriate functionality into a GIS environment. In this respect, one can consider the emerging field of spatial duration analysis. Typically, duration analysis is concerned with temporal longitudinal data, but recently has been adapted to the spatial domain since it is feasible to use any non-negative continuous variable in place of time (Pellegrini and Reader 1996; Pellegrini and Grant 1999; Rogerson, Weng and Lin 1993; Esparza and Krmenec 1996). In my research, I have applied the techniques from duration analysis to inter-event (or point) distance measurements (called spatial durations) from spatial point patterns to investigate spatial dependence, edge effects, and unobserved heterogeneity in the context of innovation diffusion and Congressional voting behavior.

However, applying such powerful statistical methods as spatial duration models is hampered by the need for the analyst to blend of tools from existing GIS software with tools in currently available statistical and econometric packages (eg., SPSS, SAS, LIMDEP). It is here that one must envision the future of spatial point pattern analysis, and explicitly taking advantage of the spatial environment provided by GIS, whereby analysts are able to perform such spatial duration analyses without continually having to cross between various computer software packages and platforms. Not only should the process be "tightly-coupled," but the functionality of existing GIS must be developed to provide the spatial analyst with the necessary, but generally atypical, data measurements involved in applying and developing the methods of spatial duration analysis. Below, a simple example from duration analysis illustrates my position.

It is fairly easy to imagine incorporating standard temporal duration models to GIS as being a relatively straightforward task, but it is entirely much more difficult to imagine providing a GIS with the functionality to make the required spatial data measurements to implement spatial duration modelling. That is, since duration analysis was developed for temporal durations, various aspects of spatial durations such as direction or weights (magnitudes) are not typically a concern of standard statistical packages. Since temporal durations, by definition, have an implied ordering and directionality, existing statistical software is restricted in its functionality for spatial analysts. Routine statistical tests like the Log-Rank test for comparing survivor functions require weighted durations for their use in spatial duration modeling, but existing statistical software does not permit such functions. Thus, this information must be measured and incorporated into spatial duration analyses by user-written routines coupling existing GIS functionality with statistical software packages. Obviously, spatial duration research extends beyond purely technical concerns such as weighting observations, to the investigation of edge effects and determining the influence of a suite of spatially-varying variables on a spatial process, all of which become increasing complex to handle without suitable spatial analytical functionality in a GIS. The link between spatial point pattern analysis and spatial duration analysis should not be overlooked since this provides a new framework within which to conceive of spatial processes and to investigate such issues as spatial censoring, spatially-varying variables and spatial heterogeneity. Spatial analysts must look towards enabling future GIS to handle the special requirements of emerging fields such as spatial duration analysis, particularly since these methods are very suitable for an assortment of research in geography, regional science, and other disciplines that study the Earth's surface. GIS will make sophisticated spatial analysis tools accessible to more users, but at the same time spatial analytical advances must be facilitated by the development of GIS technology.

**Jonathan D. Phillips**
Department of Geography, College of Geosciences, Texas A&M University, College Station, TX

## Deterministic Uncertainty and Spatial Analysis

At least some (and perhaps many) earth surface systems exhibit complex nonlinear dynamics, including dynamical instability, deterministic chaos, and self-organization due to the unstable growth of small perturbations. The spatial complexity and predictive problems associated with such behavior has been called deterministic uncertainty. The deterministic uncertainty concept differs from traditional views of spatial complexity in that spatial variability is seen as arising from the deterministic growth of perturbations or variations in initial conditions rather than (or, more likely, in addition to) stochastic forcings or apparently random environmental heterogeneity. Deterministic uncertainty differs from mainstream chaos and complexity theory (in which the term deterministic complexity is common) in that it recognizes the possibility of eventually measuring and accounting for the underlying deterministic source(s) of spatial variation. For example, a deterministic uncertainty-based view of soil variability might attribute some portion of the surface variation to variations in parent material which are unmeasurable, or which cannot be measured in sufficient detail. However, this view recognizes that improved measurement technologies might reduce uncertainty and increase predictability--i.e., the uncertainty is not necessarily irreducible.

Spatial analysis has not yet accepted the challenge posed by deterministic uncertainty, or by mainstream nonlinear dynamical systems (NDS) theory. Efforts to detect, model or assess complex behaviors in the spatial domain have largely been limited to simulation models, and have not addressed real landscapes or geographical data sets. This is largely attributable to two factors. First, many of the seminal concepts and methods of NDS theory arise from mathematical models and simple laboratory systems, and are simply not well-suited for the noisy, dirty, real world of geography and geoscience. Second, the majority of NDS work has focussed on the temporal domain. As a result, the standard methods of NDS are ill-suited to spatial data, and the standard methods of spatial analysis cannot readily distinguish deterministic complexity or uncertainty from noise.

The challenge, then, is to develop spatial analytic concepts and methods suitable for detecting and assessing deterministic uncertainty. At least three approaches are possible--the adaptation of standard NDS methods to spatial data, the adaptation of existing spatial analytic methods to complex nonlinear dynamics, or the production of new methods explicitly designed to deal with deterministic uncertainty in dirty, noisy geographical data.

GIS research is only now moving beyond the limitations of hardware and software to incorporate problem-specific spatial analysis. If GIS is to keep pace with geography and the geosciences as a whole, GIS-based spatial analysis must build upon that recent progress and confront the issues of deterministic complexity and uncertainty.

**Rhonda M. Ryznar**
Department of City and Regional Planning
University of North Carolina

My perception of the status and trends in spatial analysis come from the perspective of an urban planner. Urban planning may be distinguished within this group of UCGIS scientists in that it is a professional field. The research that urban planners carry out is intended to advance the knowledge and capabilities of the planning profession to respond to real world problems.

GIS has come to be essential to urban planning. After all, we were using overlay analysis when it really was a manual overlay process consisting of transparent mylar map sheets of various layers of site information layered on top of a base map rendering. GIS has always been viewed by the planner as a welcome technological advancement to the old-fashioned method of site planning. When GIS appeared on the scene it was positively thrilling to planners to be able to attach data to digital maps, perform calculations and derive new information to help solve old problems. Proficiency in GIS, at least in an understanding of its use and application, is essential to the planning education, and more and more planning departments are making it a regular and sometimes necessary part of the curriculum.

GIS, however, has not been very successful at making spatial analysis widely available to physical and social scientists. My impression is that most students, researchers, and professional planners do not have an adequate understanding of what spatial analysis is or what the issues of spatial analysis are in GIS. Most everyone understands the graphic nature of GIS and the value of graphical representation and the information that it can effectively convey (e.g., Tufte). Many of us also know that the locations of phenomena are important to understanding processes and many have enthusiastically embraced the computer tools that make thematic mapping and putting graphic representations in reports easier than could be done before. These are features that have helped market GIS. However, most physical and social scientists, let alone thousands of planners in towns and other government agencies, have no idea how to incorporate spatial statistics in their analyses or tap into the power that computerized locational information can add to research.

There are several reasons for this lack of knowledge or expertise as I see it:

a) Geography and the methods of spatial analysis which evolved in that discipline had all but been forgotten after the decline of geography in the United States began in the latter part of this century.

b) Spatial statistics is rarely taught in higher educational institutions. This may be changing in recent years as colleges and universities are beginning to catch on to the value of spatial analysis in research. GIS, as a technological advancement, is probably responsible for this renewed interest, but there aren't nearly enough trained professionals to teach spatial analysis.

c) GIS and spatial analysis still take a lot of time. This last one is the toughest to overcome. I experienced this as a recent Ph.D. student.

Incorporating GIS in my dissertation research added many months to my completion time and I ended up spending more time working out techniques and methods of analysis and less on planning theory. Now that I am beginning to mentor Ph.D. students myself and many show enthusiasm for incorporating GIS and spatial analysis in their research, I feel compelled to warn them of the additional effort this will necessitate. Most especially if they've never done it before!

GIS and spatial analysis is just not easy to carry out. There is a large learning curve that most physical and social scientists do not want to embark upon especially in the later stages in their careers. Some researchers have created tools and programs (e.g., Anselin, Griffth, etc.) to help make it easier and without them most of us would never be able to accomplish as much as we have. However, widely used GIS software, such as developed by ESRI, is still difficult to learn and does not have the data management and statistical and spatial analytical tools we need to do our jobs. I know this is not news to most of you, but in order to really bring spatial analysis into the classroom, and therefore, into the greater scientific and professional community, we must have better trained analysts and better tools to work with. Additionally, a comprehensive handbook of spatial analysis methods, techniques, and tools, adapted for use in GIS, is needed.

My personal research interests are in urban sustainability and the interaction between population and environment. This demands modeling processes of change which occur through space and time and incorporates data and theories from many disciplines. How to bring together a watershed model, urban development or neighborhood succession models, and the influence of transportation networks in a model of urban change through time is quite a challenge. But the notion of using locational information to unite these different data categories has been inspirational for many urban analysts. After all, all these things occur at some location. Yet there are enormous issues of scale, both temporal and spatial, and data compatibility and accuracy which must be overcome.

What tools are available to model the processes of the social, economic and natural urban environment through space and time? I have recently considered trying data mining software to seek patterns in spatial layers of data from different points in time. Data mining software is available now because of the great interest in business marketing. The software is used to identify consumer patterns and buying behavior of various socio-economic groups. This software is enormously expensive and I can't say with confidence that the methods presented are applicable for urban change analysis.

That brings me to the last point. Two of the questions asked were whether spatial analysis is being neglected by the sheer diversity of current research in GIS and will current research efforts provide an optimum environment for research in geography, regional science, and other disciplines in the coming decade? It does appear as if spatial analysis gets lost amidst the perplexity of seemingly unresolvable issues of data compatibility, scale, etc. I also think that spatial analysis is hindered by the huge amount of data that must be handled. However, while recognizing the implications of these issues on the methods and results of analysis, we must keep moving toward creating additional data management and analytical tools for the growing volumes of data we will be accessing in the future. For this reason more computer science involvement is also needed in developing future GIS. Indeed, because GIS has applications in such a variety of subject areas progress will demand multi-disciplinary team efforts.

**Daniel Z. Sui**
Department of Geography, Texas A&M University, College Station

## GIS-based Spatial Analysis and Modeling (A&M): From syntax to semantics

How to link or integrate GIS with various analytical and modeling techniques is one of my primary research interests beginning in the late 1980s. I have also been teaching a graduate level seminar entitled "GIS-based Spatial Analysis and Modeling (A&M)" at Texas A&M during the past six years. Those research and teaching experiences have given me the opportunity to scrutinize a wide-range of literature related to this topic. The goal of my position paper is to present my view on the accomplishments of making GIS a robust analytical tool and to speculate on the future directions of spatial analysis and modeling in light of the latest development. I understand that spatial analysis and modeling have meant different things to different people in different disciplines. In this paper, I do not want to make a rigorous distinction between (spatial) analysis and modeling (maybe we should do this during the workshop).

For such a daunting task, I need a high-level conceptual ladder to enable me to obtain a birds-eye view of this field. The ladder I stand on for this paper is essentially a linguistic metaphor borrowed from Casti (1994, 1997). According to Casti (1994; 1997), the essence of any analysis or modeling is a two-way mapping process: to encode certain characterizations (observable) 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 observable in the real world system. Casti 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 analysis or modeling exercises is to first find the most essential characterizations of system N, and search for the most truthful representation of these characterizations in system F. Analysis or modeling is not successful if we fail to interpret the meaning of system F in the context of system N.

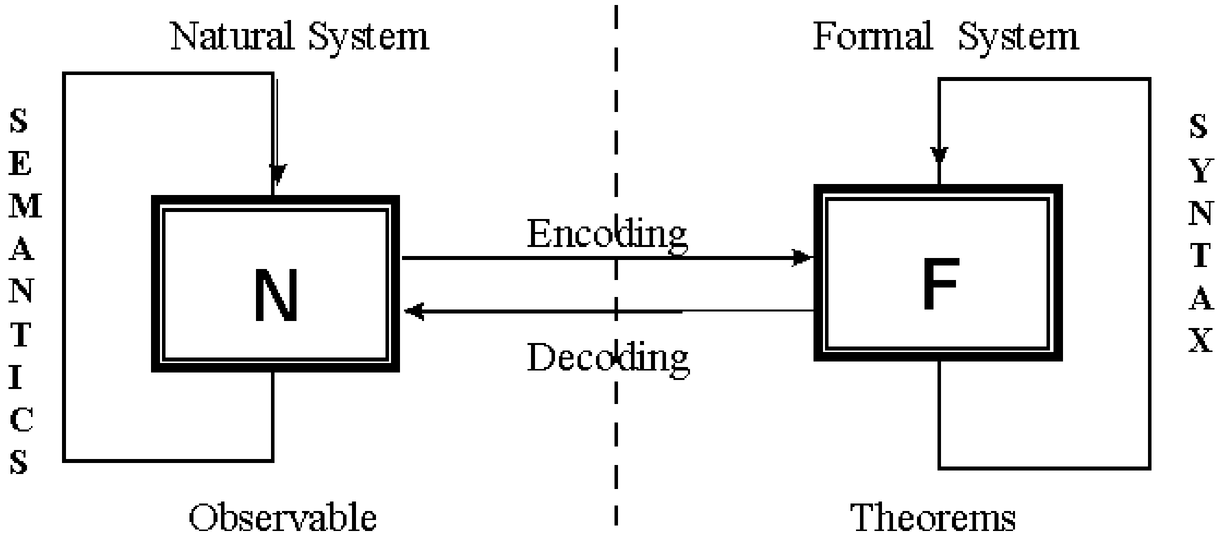


Figure 1. Syntax vs. Semantics in Spatial Analysis and Modeling
[Modified after Casti (1994)]

By applying Casti's framework to the field of GIS-based spatial analysis and modeling, I have the following observations I would like to discuss with my colleagues during the workshop: First, I believe that, in general, the efforts of trying to make GIS a robust spatial analytical tool have been less successful. The generic, syntax-driven analytical procedures, as reviewed in Bailey (1994), are still confined to academics. Neither GIS vendors nor most GIS users have developed a keen interest in those sophisticated procedures. Instead, the recent development of GIS technology seems to reflect the growing emphasis on the entire life-cycle of geographic information from data capture to storage to retrieval to visualization. To most GIS users, analysis seems to fade away as the defining GIS function. Part of the reason for this situation is caused by the fact that most GIS users (except for academic researchers) and commercial GIS software developers have a very difficult time to comprehend the semantics of these statistical procedures. Second, instead of embedding generic, syntax-driven statistic procedures in GIS, the past several years have witnessed the development of specific, semantics-based modeling procedures either as part of a commercial GIS package or as a stand-alone package capable of linking to GIS via data exchange (Wilson, 1996). These modeling procedures are generally tied to a very specific domain that has a set of commonly accepted models in practice. Products such as RiverCAD, HEC-RAS, TransCAD, and TransPlan etc. are some of the examples of this trend. The market demands for these specific modeling functions /packages (not the generic spatial statistic procedures) seem to continue to grow, but the development of these specific modeling tool boxes is most likely to be in the hands of GIS users and researchers, either using the built-in GIS script languages such as Avenue or MapBasic or other high-level languages such as VisualBasic, C++ etc. Current efforts toward the interoperable GIS will greatly facilitate the implementation of various models in the GIS context. The main role of GIS is essentially to provide modelers a consistent digital representation for them to implement their specific models. Third, as for the future of GIS-based spatial analysis and modeling, this paper favors a semantics-driven approach with emphasis on contextual meanings. Instead of looking for generic, spatial statistic procedures detached from specific contexts, this semantics-driven approach will serve us well not only in practice (to meet the growing demands of socio-economic and environmental modeling using GIS) but also in research. The semantics-driven approach will inevitably lead us to address questions raised at both ontological and epistemological levels. By shifting from a syntax-driven to a semantics-driven approach, we can better address those critics of GIS from social theorists as well as scientists from specific disciplines. Our answers to these ontological and epistemological questions will determine, to a large extent, what kind of spatial analysis and modeling practice we will conduct tomorrow.

I will have a more polished paper written before the workshop and it will be circulated among the participants during the workshop.

**References Cited**

Bailey, T.C., 1994. *A Review of Statistical Spatial Analysis in Geographical Information Systems*. In S. Fotheringham and P.
Rogerson (eds.), *Spatial Analysis and GIS*, pp.13-44

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.)

Wilson, 1996. *GIS-Based Land Surface/Subsurface Modeling: New Potential for New Models? Proceedings of the Third International Workshop on Environmental Modeling*. http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/wilson_john/wilson.html

**Lance A. Waller**
Department of Biostatistics, Rollins School of Public Health, Emory University

## SOME THOUGHTS ON GIS AS A SPATIAL ANALYSIS TOOL

As we are all influenced (and in some sense limited) by our respective areas of training, let me note that my interests in spatial analysis and GIS arise primarily from spatial statistics as applied to the analysis of public health data. From this context, I briefly describe two spatial analysis issues of which I believe cross between many application areas seeking to implement GIS as a research tool.

First, an often-touted strength of GIS involves the linking of spatially referenced data sets, many times collected by different agencies for different purposes. Assessments of data quality and accurate analysis (and visualization) of data uncertainty are critical to attaching scientific inference to data linked and displayed in a GIS. The accurate use of GIS in spatial analysis requires development of tools for addressing uncertainty (both statistical and deterministic) within the GIS environment. Otherwise, GIS offers many advances in creative data display and management, but actual analysis of data primarily occurs outside of the GIS (either in the head of the viewer of GIS output, or in specialized software packages). In short, without accurate presentation of uncertainty, associations between variables displayed in a good map of bad data often appear more believable than those displayed in a bad map of good data, and there is sizable potential for misinterpretation.

Second, many application areas focus on the analysis of observational rather than experimental data. Different aspects of key issues in the analysis of observational data surface in different application areas. For example, the "modifiable areal unit problem" of geography has aspects in common with sociology's and epidemiology's "ecologic fallacy" of assigning associations observed in aggregate to individuals. Also, the latent variables of econometrics and the notion of unmeasured confounding in epidemiology are differently named manifestations of the same problem. While these issues are not necessarily synonyms, they nonetheless reflect different facets of deeper issues underlying all analyses of observational data. As a result, there is need for more interdisciplinary collaborations providing fundamental advances in spatial analysis of observational data without "reinventing (or renaming) the wheel".

Advances in the utility of GIS and spatial analysis in any of a variety of application areas similarly would profit from interdisciplinary developments. GIS advocates the introduction of spatial thinking into application areas, however I also believe GIS can benefit from the introduction of application area thinking. For example, the field of epidemiology can benefit from the introduction of spatial analysis techniques, but not at the sacrifice of well-developed concepts such as confounding, causation, and the ecologic fallacy. The underlying goal in introducing GIS as an analysis tool should not be only to enable novel analyses, but to enable better analyses. In some cases this will involve construction of new analytic techniques enabled by GIS, but in others the true utility of GIS lies in increased efficiency in design and implementation of established approaches. Determining which is most appropriate requires creative insight from GIS developers, spatial analysts, and application area experts.

**Elizabeth A. Wentz**
Department of Geography, Arizona State University, Tempe

Geographical analysis was one of the motivating forces behind the beginnings of geographic information systems (GIS). It is widely recognized, however, that the development and utilization of analytical capabilities within GIS has never realized its potential as a tool for spatial analysis. Nevertheless, there is much to be gained for incorporating spatial analysis with GIS. Spatial data are more abundant than ever and analytical approaches are needed to understand the geographic processes that these data describe. Additionally, geographic problems are being defined in more complex manners than before and they address issues that are relevant to society and everyday life. Increasing the ability to examine these problems with analytical approaches has been attempted through interoperability of spatial technologies (GIS, remote sensing), investigation to better understand the properties of spatial data (MAUP, scale, error), and investigation into new methods of dealing with large data volumes (visualization and data mining). The statement I present here summaries my views on the reasons I believe spatial analysis is not yet integral to GIS, describes approaches to solving the problem that I believe do not work, and highlights what I believe is needed in order to improve the overall analytical capabilities of GIS.

Although analysis is an integral to GIS by definition, it has not been fully integrated into functioning systems to any significant degree for several reasons. First, early research and development efforts were put toward data input and storage techniques. This put analysis issues in the background in terms of research and development priorities. A second reason why analysis is still not an integral component of GIS is that there is no perceived need for analysis within a large portion of the GIS user community. Many GIS users benefit simply from the storage capabilities of GIS for the inventory of spatial data. These applications, as they have currently defined them, do not require the use of analytical functions. Other users who are adopting GIS are accustomed to obtaining analysis capabilities from other sources, such as environmental models and spatial statistical analysis (e.g., hydrologic, transport, and groundwater models). For them, GIS represents a means of data management and efficient access rather than an integral part of the analytical process. The problem of no perceived need for analysis by the users is exacerbated because the software industry is market driven. The software industry is working toward improving the access to spatial data through “user friendly” and “easily accessible” tools and yet increase “power” and “flexibility” of the overall software. The concept of an “easy” yet “powerful” system is potentially contradictory.

One of the overriding trends to solving the problem of incorporating spatial analysis in GIS is based on developing links between software packages. The benefits of such a linkage are clear. The result of linked software would be an integrated environment where research effort could be applied to solving a particular substantive problem rather than managing the technology. This approach to improving spatial analysis in GIS offers the advantages that the techniques exist, are known to the users, and the algorithms and code exist. However, this approach as the basis for a research agenda to improve the analytical capabilities of GIS is limited. Developing an integrated environment is simply a technical one and does not improve our ability to solve problems beyond the pre-computer methods to address the complexity of real-world spatial issues.

In general, I believe spatial analysts and GIS specialists who wish to integrate spatial analysis techniques into GIS are taking a research focus that is too narrow. My goal as a researcher to expand the analytical capabilities of GIS by broadening the perception of spatial analysis in GIS to open up new research opportunities. The avenues by which to accomplish this expansion are several.

First, there remains reluctance to identify modeling as a component to spatial analysis. Modeling, although often specific to one application context, represents the culmination of many spatial analysis techniques. Limiting research to analytical processes just prior to this step is shortsighted and does not serve to broaden the capabilities of GIS. Research examining the related aspects of spatial statistics and modeling should be examined to define a comprehensive framework for spatial data analysis.

Secondly, spatial analysts rarely look at what GIS can offer spatial analysis beyond basic data storage and manipulation. There is the common perception that GIS can gain significantly by integrating more analysis (again, a view that I believe is a technical issue). Rarely do we examine the ways in which research in spatial analysis can benefit from the strengths in GIS. Needed is a turn toward examining how GIS can benefit spatial analysis, such as definitions of geometric properties, topology, and visualization.

Thirdly, the computer is not yet viewed as an alternative medium to the paper map for representing spatial data. This view should be reflected in the assumptions associated with defining spatial data and subsequently the form of the analysis that are performed on these new data types. With the computer as an alternative medium for representing spatial data, we are not limited to describing spatial phenomena with points, lines, and areas. Definition by objects, for example, allows us to build spatial representations that include both the data and the processes. Needed is research toward defining objects that can describe complex phenomena such as qualitative data and spatio-temporal data. Questions regarding how alternative forms of data representation influence and change the types of analysis performed can then be examined.

Finally, spatial analyst are not examining what is happening with regards the increasing rate of data processing, which is resulting in the definition of near real-time problems. The rate of information processing is increasing due to the expansion of the Internet, direct data capture through remote sensing and GPS, and improved hardware. There is growing need for developing concepts to go along with the data processing such as real-time GIS and virtual laboratory. The role of role of spatial analysis in these environments needs to be evaluated. Applications are being built to use the Internet and available data in contexts such as:

- use of real-time GIS for emergency evacuation routes
- electric companies assessing the cause of power outages and dispatching repair crew
- palm-top GPS tracking with GIS data handling capabilities
- real-time GIS for storm tracking and analysis
- simulation laboratories for pollution modeling

These types of applications are growing and the role of spatial analysis in them is virtually undefined.

In summary, geographic problems are being defined to include qualitative data, spatio-temporal relationships, and the expansion of these data into multiple scales. Research in spatial analysis and GIS must recognize the kinds of problems that are being defined and build tools to help answer them. Spatial analysts need to untie themselves from the traditional techniques, and move GIS and spatial analysis beyond replicating the paper map.