ARTHUR GETIS

CHAPTER 5

THE ROLE OF GEOGRAPHIC INFORMATION SCIENCE IN APPLIED GEOGRAPHY

ABSTRACT

Applied geography has undergone remarkable changes in the last 20 years. Powerful new technologies have emerged that greatly improve the ability to collect, store, manage, view, analyze, and utilize information regarding the critical issues of our time. These technologies include geographic information systems (GIS), global positioning systems (GPS), satellite-base remote sensing, and a great variety of remarkable software that allows for the analysis of the compelling problems. The issues include globalization, global warming, pollution, security, crime, public health, transportation, energy supplies, and population growth. Geographic Information Science (GIScience) has given rise to an essentially multidisciplinary approach to applied problems. No single person is expert in all of these areas. It is necessary to emphasize coordination and collaboration and to find the bridges that reduce the barriers between disciplines. In this chapter we briefly discuss the new technologies and the way in which they are being used to solve the critical issues. We then make suggestions for an applied geography future vis-à-vis the geographic information sciences.

INTRODUCTION

In the early 1980s, the world of applied geography began to undergo a radical change. The desktop workstation made information storage, retrieval, manipulation, and processing faster and offered greater capacity than geographic problem solvers had ever had before. By the mid-1980s, geographic information systems (GIS) software added the important mapping dimension to data display and data organization. One could use software such as SURFER in the late 1960s and 70s, but the great power and ease of the personal computer in the 1990s changed the ground rules. One no longer had to gather a huge packet of Hollerith cards punched on an IBM 029 and troop across the campus to the computer center for processing. Now, in the mid-2000s, not only is it routine activity to use GIS software, but the nature of GIS itself is changing dramatically. If productivity is any measure of technological gain, it would be fair to say that spatial data analysis is perhaps 10,000 times more powerful in 2003 than in 1983. Moore's Law holds here, that is, about every year or so the speed of computer processors doubles. Data storage for inexpensive personal computers has gone from 64 kilobytes in the

A. Bailly and L. J. Gibson (eds.), Applied Geography, 95–111.

mid-80s to 60 megabytes in the mid-90s to 60 gigabytes today. Soon, terabytes will be the common way to measure computer capacity.

Other chapters in this volume cover traditional areas of concern of applied geography. These subjects have changes only modestly during this period of great technological change. Interest continues in commercial and industrial location, transportation, crime, public facility location, disease distributions, and environmental concerns. But instead of processing data and drawing maps by hand, now we depend nearly entirely on computers to help us in our research endeavors. The traditional research process – problem to hypothesis to model to data gathering to test to decision – has been altered. Because of the power and flexibility of computers, the process has changed appreciably to include: problems, large scale data gathering, data manipulation including data visualization, model building, hypotheses generation, simulation, model validation and testing. The prescribed order of research is much more idiosyncratic now. While much can be accomplished with the new technology, more data are needed in order to satisfy the appetite of the more demanding research process.

National and international geographic information science

In the mid-1990s, a new organization was established in the United States, the University Consortium for Geographic Information Sciences (UCGIS). This was the first formal recognition that GIS is but a subset of the broader field of *GIScience*. Similar organizations have been established in Europe (AGILE) and Japan (GIS Association of Japan). Originally, most GIS practitioners were geographers, but with the burgeoning of the field and the need for more and more sophisticated technology, engineers, computer scientists, and practitioners in fields that have recognized the value of GIS have attached their interests to the GIScience bandwagon. These related interests cover such areas as spatial data acquisition and integration, distributed computing, cognition of geographic information, interoperability of geographic information among computer platforms, and remote sensing. As a recognition of this broadening field, the term GIScience is now being used to express many of these interests, not to the exclusion of GIS but in conjunction with GIS. This has stimulated a multidisciplinary approach to the solution of applied geographic problems. In fact, it is becoming more and more difficult for individual researchers to engage in a research project without some sort of interdisciplinary collaboration. For example, the solution of a facility location problem might require the cooperation of a modeler, GIS specialists in data base management and graphics design, a remote sensing expert, and a spatial statistician.

Several National Science Foundation supported centers have been developed over the last several years. The purpose of these institutes is to inform interested researchers, teachers, and practitioners of the work being done in the spatial sciences and to instruct those new to the field in the techniques of spatial science. These organizations include the Center for Spatially Integrated Social Science (CSISS) and Spatial Perspectives on Analysis for Curriculum Enhancement (SPACE). In this chapter, we outline what GIScience means, how it is used in applied geography, and give examples of its use for a variety of problems often addressed in applied geography.

THE CONTENT OF GISCIENCE

In the last several years the UCGIS has begun to create model curricula that when completed will represent the subject matter of GIScience. The model curricula are designed to satisfy the needs of educators who want to prepare students for the use of geographic information and technology in their post-graduate work. The purpose of having curricula rather than a single curriculum is to afford students from a wide variety of disciplines and emphases the opportunity to follow relevant paths toward a useful knowledge of GIScience. For those working in applied geography, nearly all of the subject areas covered in the model curricula will be of interest. One can convincingly argue that a working knowledge of GIScience is mandatory in order to do research in applied geography. The major elements of the model curricula (as of mid-2003) are called knowledge areas (see www.UCGIS.org). These represent the content of GIScience. The following brief sketches define the various aspects of the knowledge areas.

Conceptualization of space. The nature of space and time is the context for earth-related phenomena. This includes different notions of space and time for differing applications and different disciplines. It includes the means to understand scale, pattern, location, and region, and forms the basis for dealing with the entirety of GIScience.

Formalizing spatial conceptions. Here we cast our conceptual view of space and time into a specific, logical organizational structure from a given application perspective. Formalization incorporates a set of specifications to measure, reference, and locate spatial and spatial-temporal conceptualizations. Examples include modes of measurement, coordinate systems, map projections, spatial relationships, topology, object and object-type categories, diffusion, and network flows.

Spatial data models and data structures. This is the representation of formalized spatial reality through data models, and the translation of these data models into data structures that are capable of being implemented within a computational environment. Examples of spatial data model types are discrete (object-based), continuous (location-based), dynamic, and probabilistic. Data structures represent the operational implementation of data models within a computational environment.

Design aspects of GIScience and technology. Analytic model design incorporates methods for developing effective mathematical models of spatial and spatial-temporal situations and processes. *System design* addresses the manner in which existing GIScience concepts and technology are matched with the requirements to create solutions to operational spatial problems. *Spatial Database Design* concerns the optimal organization of spatial data.

Spatial data acquisition, sources, and standards. Data acquisition is required for the development of fundamental data layers within a GIS. These data may have spatial, temporal, and attribute (descriptive) components. Examples of primary data sources include surveying, remote sensing (air photography, satellite imaging), the global positioning system (GPS), work logs (e.g., police accident reports), and surveys. Secondary spatial or spatial–temporal data can be acquired from digitized and scanned analog maps as well as from sources such as governmental agencies. Data standards for spatial data, images, and metadata exist to document data quality, lineage, and appropriate use.

Spatial data manipulation. This includes the ways in which one transforms spatial data into formats that facilitate subsequent analysis. Examples of data manipulation include vector-to-raster conversion, line generalization, attribute aggregation, projection transformation, and transaction management.

Exploratory spatial data analysis. Such analysis includes operations whose objectives are to derive summary descriptions of data, evoke insights about characteristics of data, contribute to the development of research hypotheses, and lead to the derivation of analytical results. This is also called *data driven* analysis.

Confirmatory spatial data analysis. This includes the techniques used to create and test spatial and spatial-temporal process models. It may also be called *model-driven analysis.* This area is tied directly to specialized problems studied in the social, behavioral, and physical sciences. For example, an applied environmental geographer would want to learn the confirmatory analytic procedures that are particularly well suited for modeling a spatial environmental process such as air or water pollution.

Computational geography. This is the application of computationally intensive approaches to the study of the geosciences. The focus of geocomputation is a variety of methods designed to model and analyze a range of highly complex, often non—deterministic, non—linear problems. These include such areas as neurocomputing, fuzzy sets, and genetic algorithms.

Cartography and visualization. This is the creation and effective interpretation of graphic representations of spatial data sets and of the results of spatial analysis activities.

Professional, social, legal aspects of GIScience and technology. This covers the nature and extent of (a) property rights in the spatial information itself, (b) the use of spatial information for land management and other decisions by both public and private actors, and (c) the distribution of the spatial information. Administrative practices are concerned with the nature and extent of how organizations and agencies are organized and made into effective operations that manage the information and its application.

METHODS OF SPATIAL ANALYSIS FROM A GISCIENCE PERSPECTIVE

Because of technological advances, researchers in the social and physical sciences seeking empirical verification of their models are faced with larger data sets made up of more detailed spatial information than has been the case in the recent past. Among the types of expanding data sets are: detailed consumer surveys (household data), the point location of crime and other social variables, detailed land cover information (remotely sensed data), environmental indicators (air and water quality data), the geo-referenced selling price of houses (real estate data), traffic flow (origination and destination data), consumer demand (individual choice data), and migration (individual household movement data). As the new technology evolves, larger and larger data sets will contain information on smaller and smaller spatial units. Small unit data, such as census block data, are currently available in easy to use graphic form. Data pinpointing 911 calls are an example of available detailed spatial information. Of course, in due time detailed historical data of this nature will be available for study. Currently, *The National Historical Geographic Information System* project is preparing for analysis socioeconomic data from United States censuses dating back to 1790. Maps are being prepared that will allow one to follow changes in geographic areas over time.

Coming to grips with the new technology is not a straightforward undertaking for applied geographers who are used to modest sample sizes, aggregate data, and models that require assumptions about data. Nonetheless, the new technology should be welcome for a number of reasons.

- The need to test theory at the scale that the theory requires tends to favor larger scales of analysis over smaller scales. Most social, economic, and environmental theory is based on individual rather than group behavior.
- Applied geographers usually opt for larger rather than smaller data sets, not only so that they may obtain more reliable estimates of model parameters, but to reserve data for sensitivity analyses.
- Increasingly, applied geographers favor analytical schemes in which the strict rules of statistical theory can be relaxed. Fast computers capable of handling large sets of data are ideal for devising simulation schemes that are less dependent on the complications of statistical rigor.
- Exploratory data analysis can become a highly productive prelude to confirmatory data analysis in applied geography.

Data preparation and GIScience

Of great concern to applied geographers is the preparation of data for use in model building and eventually in the confirmatory analysis that model building implies. Small spatial unit data suffer from a variety of inherent problems that make them generally unfit for confirmatory analysis. These problems revolve around the assumptions made about the distribution of data that are to be used to test models. The main issues are:

- Finding the appropriate scale of analysis. A GIScience approach, if the data allow, can be used for the study of a problem in multiple scales.
- Recognizing the degree of stationarity of spatially distributed variables. A GIScience approach allows for the kernel estimation of the density of phenomena over the study surface.
- Identifying the degree and characteristics of any spatial association in the data. Spatial statistics now embedded in GIS software provides measures of the degree of spatial autocorrelation in data.

Although exploratory spatial data analysis (ESDA) techniques have been in existence for a long time, GIScience technology allows for the manipulation of large amounts of data quickly. Models can be developed by means of identifying particularly constructive aspects discovered in an ESDA. In the process of ESDA, models become clearer, data are prepared for analysis, and problem solving can be carried out effectively. This sequence does not obviate the need for looping backwards to ESDA to stimulate further thought on the nature of the proposed model.

Data analysis

The types of statistical methods used in applied geography are a function of both the nature of the problems to be solved and the availability of the new faster technologies. We list below four general areas that are particularly useful for applied geography. Each is described in terms of the kinds of problems being solved, their general formulation, and their usefulness within the GIS community of analysts. Many new techniques have not yet been tied to hypothesis guided inquiry. Such areas of inquiry as spatial neural nets, spatial fuzzy sets, and simulated annealing are just now being developed; they are not discussed here.

Pattern analysis

Popular in the 1960s was point pattern analysis based on the spatially homogeneous Poisson process. It was common to find a researcher working at a light table making measures from pencil-numbered points to the first nearest neighbor of each point. Now, with the use of digitized georeferenced data, we are easily able to take measurements from all points to all other points. In addition, measurements of line segments, distances between line intersections, areas, and characteristics of areas such as perimeter length, neighboring areas, and so on, are basic measuring rods within many GIS.

Pattern analysis in the spatial sciences grew out of a hypothesis testing tradition, not from the extensive pattern recognition literature. Nearest neighbor work continues today, but the work of Clark and Evans (1954) has now been modified for the sake of unbiasedness to take into account the length of the perimeter of study areas and the distance to study area boundaries (refined nearest neighbor analysis) (Boots and Getis 1988). Today, nearest neighbor measurements are a fundamental function of many GIS.

Perhaps the most important developments in recent years are the application of K-function analysis to the study of point patterns and the use of Voronoi polygons for the study of spatial tessellations (Okabe et al. 1992). In addition, fractals study is a promising area for pattern analysis (Batty and Longley 1994). A brief outline of the idea behind the use of K-function analysis and some of the successful applications to spatial phenomena follow.

The K-function is the ratio of the sum of all pairs of points within a pre-specified distance d of all points to the sum of all pairs of points regardless of distance (Ripley 1981). The function is adjusted to identify distances that are closer to the boundary of the study area than to d. The K-function takes into account the need to stabilize variance, and Getis generalized the formula to include the weighting of points, such that the sum of pairs of points became the sum of the multiples of the weights associated with each member of a pair of points (1984). Diggle (1983) has successfully exploited this formulation to show many new features of patterns. For example, not only can one easily show the difference between an existing pattern and a random pattern, but also one can develop theoretical expectations for other than random patterns. In addition, patterns divided into different point types (marked patterns) can be studied easily. For testing purposes, an envelope of possible outcomes under the hypothesis of say, randomness, is usually constructed by means of a Monte Carlo simulation.

Although studies of the spatial distribution of vegetation dominate the empirical literature of K-function analysis (Diggle 1983), the method has been used for the study of population distribution and disease distribution. Bailey and Gatrell (1995) showed that the K-function can be used as an indicator of time-space clustering. That is, one simultaneously finds pairs of points separated by designated units of time and distances in space. This approach is particularly useful for identifying disease clustering over time.

Spatial association

Finding the degree of spatial association (autocorrelation) among data representing related locations is fundamental to the statistical analysis of dependence and heterogeneity in spatial patterns. A fundamental feature of most GIS packages is the Moran and Geary measures of spatial autocorrelation. Moran's statistic, very much like Pearson's product moment correlation coefficient, is based on the covariance among designated associated locations, while Geary's takes into account numerical differences between associated locations. The tests are particularly useful on the mapped residuals of an ordinary least square regression analysis. Statistically significant spatial autocorrelation implies that the regression model is not properly specified and that one or more new variables should be entered into the regression model. Research in this area has emphasized the distribution characteristics of the statistics under varying spatial resolutions and spatial weights matrices (Anselin and Rey).

Mantel (1967) and Hubert (1979) have shown that statistics of this nature are special cases of a general formulation, gamma, that is defined by a matrix representing possible location associations (the spatial weights matrix) among all points multiplied by a matrix representing some specified non-spatial association among the points. The non-spatial association may be an economic, social, or other relationship. When the elements of these matrices are similar, high positive autocorrelation obtains. Gamma describes spatial association based on co-variances (Moran's statistic, I), or subtraction (Geary's statistic, c), or addition (Getis and Ord's statistic, G). These statistics are global insofar as all measurements between locations are taken into account simultaneously.

When the spatial weights matrix is a column vector, gamma becomes local; that is, association is sought between a single point and all other points (I_i, c_i, G_i). Research on local statistics has been especially active recently since these types of statistics lend themselves to kernel-type analyses in a GIS where data sets are large (Anselin 1994, Getis and Ord 1992, Ord and Getis 1995). Local statistics have been used to classify remotely sensed data, and show associations between neighborhood crime rates and the conflict propensities of countries.

Geostatistics

The variogram, discussed earlier in this book, plays a useful role as the function that describes spatial dependence for a regional (georeferenced) variable. The term intrinsic stationarity is used to describe the natural increase in variance between observations of a regional variable as distance increases from each observation. The semivariance, a measure of the variance as distance increases from all points or areas (blocks), eventually reaches a value equal to the variance for the entire array of data locations, regardless of distance. Clearly, at zero distance from a point, the semivariance is also zero, but the semivariance increases until, at a distance called the range and a semivariance value called the sill, the semivariance is equal to the variance. The function describing the semivariance is usually spherical, exponential, or Gaussian. The variogram is essential for kriging, which is a technique for estimating the value of a regional variable from adjacent values while considering the dependence expressed in the variogram. There are many kinds of kriging, each designed to give the highest possible confidence to the estimation of a variable at non-data locations. If there is no bias in the variogram, and all required assumptions are met, the kriged values, as opposed to trend surface, TIN, or other estimation devices, will be optimal. A large literature has developed in geostatistics. The definitive text by Cressie (1991) details many instances where the geostatistical approach has proved helpful. These include studies of soil-water tension, wheat yields, acid deposition, and sudden infant death syndrome. The variogram has now been introduced into several GIS, and several programs that can be interfaced with GIS are available to help construct variograms and apply the kriging process. The geographic literature in this area is building rapidly.

Spatial econometrics

The fundamental work in this area can be traced to Paelinck in 1966 (Paelinck and Klaassen 1979). Anselin has made spatial econometrics accessible to a wide audience with his text (1988) and now his SpaceStat software. In addition texts by Haining (1990) and Griffith (1988) have helped to widen the appeal of these methods in geography. The approach is as Anselin says, 'model driven;' that is, the focus is on regression parameter estimation, model specification, and testing when spatial effects are present. Regression models constitute the leading approach for the study of economic and social phenomena. The assumptions required for the basic linear regression model, however, do not satisfy the needs of spatial regression models; they must take into account spatial dependence and/or spatial heterogeneity. Spatial dependence occurs when there is a relationship between observations on one or more variables at one point in space with those at another point in space, while spatial heterogeneity results from data that are not homogeneous, for example, population by areas which vary considerably by size and shape.

A variety of spatial autoregressive models have been developed that include one or more spatial weight matrices that describe the many spatial associations in the data. The models include either a single general stochastic autocorrelation parameter, a series of autocorrelation parameters; one for each independent variable conditioned by spatial effects (dependency or heterogeneity), an error term autocorrelation parameter, or some combination of these. Parameter estimation procedures can be complex. The usual approach is to use diagnostic statistics to test for dependence and/or heteroscedasticity among these spatially weighted variables or error term. Fortunately, SpaceStat, designed for the exploration and testing of spatial autoregressive models, is sufficiently user friendly to allow for the development of final autoregressive models.

Several other approaches have been taken to specify the influence of spatial effects in a regression model environment. Casetti's expansion method (Jones and Casetti 1992) is designed to increase the number of variables in a regression model to take into account secondary, but influential spatial variables, such as the x, y coordinates of georeferenced variables. This approach uses the parameters of the expansion variables as the indicators of the spatial effects.

In another development, Getis (1995), Griffith (1996), and (Getis and Griffith 2002) suggests transforming the spatially autocorrelated model into one without spatial autocorrelation embedded within it. By filtering out the spatial autocorrelation, the ordinary least squares model can be estimated and evaluated using R^2. The filtered spatial components are reentered into the regression equation as separate spatial variables.

GEOGRAPHIC INFORMATION SCIENCE IN APPLIED RESEARCH

Of the thousands of examples of the use of GIS in applied research, I have selected a few from each of a number of fields in order to illustrate the rich and diverse nature of GIScience. Most important, however, from an applied geography perspective is the problem that motivates the research. Our point below is that GIScience represents a superb pathway for the solution of applied geographic problems. Thus, in each of the sketches we identify the problem, then point out the role of GIScience for its solution. Every issue of *ArcNews*, a monthly publication of the ESRI company, describes numerous examples of 'GIS in Action' from a variety of fields. Many of the discussions give a website for readers who want further information about a particular subject. In addition, Goodchild and Janelle (2003), present detailed examples of research examples in GIScience and Longley, et al. (1999) provide extensive surveys of practical applications of GIS.

GIScience and disaster response

One of the strongest examples of the use of GIScience took place in the days immediately following the disaster of September 11, 2001. The problem was to help design an evacuation (for debris) and rescue equipment transportation network. With smoke billowing from the site of the attack, and the mayor's planning office destroyed, infrared images were gathered so that one could 'see' through the smoke. A planning system was devised using detailed maps that allowed large equipment to get into and out of the World Trade Center site.

Because natural and human-induced disasters such as tornadoes, hurricanes, earthquakes, wildfires, and floods usually occur suddenly, threatening people and structures, they create chaos and panic (Greene 2002). Increasingly, GIS technology is being used to help communities prepare for disasters, create response plans, track and assess damage resulting from a disaster, and coordinate workers in the field (Briggs et al. 2002). Whether the disaster is a wildfire in Arizona or a tornado in Oklahoma, maps produced using GIS have been invaluable in tasks such as locating houses, identifying property ownership, helping responders decide where to send field crews or rescue workers, generating evacuation routes, and siting field hospitals or paramedic bases. Using GIS, the turnaround time is so short that maps can be updated quickly and given to crews in the field. In the wake of the disaster, maps have been used to locate damaged structures, make property damage assessments, and prepare for debris removal.

GIScience and epidemiology

In what way can infectious disease diffusion be controlled? Here, as in all GIScience solutions, a significant resource is the spatial data management system (SDMS).

Accurately mapped detailed information is needed for the location of the incidence of disease over time, entomological risk factors, and healthcare facilities. These data are linked together in a relational database. Such diseases as malaria, West Nile virus, SARS, AIDS, dengue, and Ebola result from the spread or diffusion of one or more viruses. For dengue, the mosquito species *Aedes aegypti* is the vector that carries dengue viruses from infected people to susceptibles, usually children. The data are statistically analyzed to determine the degree of spatial congruence between infecteds and susceptibles. Recommendations for disease mitigation result.

GIScience and public health

How can the delivery of healthcare service be improved? In a GIScience environment, the solution to this problem requires the detailed management of a large amount of georeferenced socioeconomic data (Cromley and McLafferty 2002). The data are manipulated into healthcare regions of varying bases dependent on such items as physicians services, clinics for a variety of healthcare needs, and transportation networks. The regions are compared and contrasted and evaluated with regard to available financial and human resources.

GIScience and telecommunications

To what extent and in what ways does electronic mail substitute for international telephone flows? In this case, the SDMS includes data on international telephone flows, their origins and destinations, and routes. These data in conjunction with variables such as Internet access, language, tourism, and income are manipulated to identify associations, spatial and non-spatial. Such geographic variables as time zones and nearness are studied using spatial software.

GIScience and demographic research

More and more demographers are finding that the spatial component of their work has much to offer in their explanations of age, sex, birth and death rate differences. For example, the question has arisen about the influence of nearby areas, for example, the neighborhood, that may transcend more theoretically plausible explanations for birth rate decline. This kind of work usually requires extensive georeferenced databases.

In a particular instance, that is, for a study of rapid changes in patterns of fertility in the developing world, it was necessary to identify regional differences among demographic variables. For national planning purposes, the research is geared to finding correlates of fertility changes. Data from census units provide information on age, number of children, socio-economic and neighborhood characteristics. In addition, remotely sensed data may be used as a surrogate to identify the location of regions of urbanization. These are entered into a relational database keyed to the centroid of each census unit. A regression model is used to explain the level of fertility vis-à-vis patterns of marriage and female education.

GIScience and economic research

For many years economists paid scant attention to the location of economic activity. With the arrival of the field of regional science in the 1950s and its enhancement by GIS in more recent years, considerable emphasis is now being given to locational phenomena Bateman et al. 2003). As an example, economists have given new meaning to the convergence hypothesis. That is, economic theory would predict that in a free choice setting regionally disparate incomes would tend to converge to the same level. The study of such a phenomenon requires an understanding of the nuances and influences of the structure of the underlying data. For example, to what extent does aggregated data (say data aggregated from census blocks into census tracts) distort the convergence process? Also, how does spatially correlated data influence results of such studies. This type of analysis is representative of what is called a spatially autoregressive process. Software packages make it possible to model the spatial effects of many types of economic phenomena.

Applied geographers consider the study of the location of economic activity as the fundamental building block of their area of study. A large literature has helped in the modeling of efficient locations for such things as retail outlets and industrial plants. In the past, maps were used mainly to display the location of inputs, outputs, and resulting locations. Now, through the visualization mechanisms of GIScience, one can map, chart, graph, and otherwise "see" the interconnections between production and consumption. Better solutions are available mainly because of the new ability to manipulate data in a myriad number of ways (Birkin 1996).

GIScience and urban and transportation research

Fundamental to urban research is the determination of what may be called optimal land use systems. Optimality is usually defined in terms of the immediate objectives of a study; for example, an optimal transportation system may be one that minimizes delays due to congestion. Again, any success in land use modeling is usually traceable to a detailed, extensive, and well-structured data set. For transportation research this usually means that origin-destination data are available as well as data on the use of various transport modes at different times of day (Lang 1999). A significant challenge is to design such a data system that can be used to study various transportation scenarios. With projections of increasing population, for example, plans can be developed that will shed light on urban congestion.

GIScience and sociological research

The degree of segregation or social exclusion has been a recurrent research theme in sociology and geography. After acceptable and rigorous definitions of what is meant by segregation and exclusion, the sociologist will depend on detailed, usually urban, data to identify clusters of exclusion and inclusion. Spatial statistics and geostatistics can be used to show changing structures of segregation over time.

GIScience and criminal justice

Given limited budgets, police departments must constantly find ways to allocate resources in the most effective ways. The local citizenry is ever alert in making recommendations as to how crime might be reduced, but it is the sophisticated SDMS that forms the foundation of effective analyses and mitigation. There are many examples of the use of GIS to identify the location of 'hot spots' of criminal activity so that police resources can be redirected to them at a moment's notice. The hot spots are of two varieties: the first are locations of elevated levels of criminal activity. More often than not, these are expected centers of crime based on population density, economic, and age data. The second type of hot spot analysis identifies crime levels statistically significantly greater than expected. The hot spots become the locations of out-of-the-ordinary criminal activity, thus allowing for redeployment of resources. Spatially based software packages are now readily available to crime solvers for these types of analyses.

GIScience and political science

Software exists that makes it possible to redraw legislative districts under a variety of criteria and scenarios although, in the final analysis, politicians usually determine where district boundaries are drawn. Sometimes, of course, courts overturn a particular districting "solution." GIS speeds up a very slow, mandatory process.

One type of approach to redistricting combines apportionment equality subject to compactness and contiguity constraints. More usual, however, is the approact that maximizes compactness while satisfying apportionament conditions such as a maximum permitted deviation from a quota. Most of this work is based on data from the smallest possible census zones (blocks and tracts). A particularly useful aspect of this is the opportunity to evaluate already existing legislative zones or zones being proposed by various legislative or lay groups.

GIScience and the environment

No subject area is better suited for study by means of GIScience than the environment. Numerous environmental problems have been and are currently being studied in a GIS setting. These subjects include protection of endangered species, reduction of air and water pollution, the evaluation of fire hazards, wildlife management, optimum habitat locations, natural resource use, landscape conservation, and the monitoring of environmental variables. These subject areas require extensive data collection and data organization. Mapping in an exploratory setting aids in the development of hypotheses. Models based on environmental relationships can be evaluated using a series of map-based statistical tests (Skidmore and Prins 2002).

GIScience, government, and public policy

Local and regional governments are constantly faced with the problem of finding efficient locations for services rendered. One recurrent problem is that of school locations and school service areas. Such techniques as the location allocation procedure of p-median have now been incorporated into GIS software so that alternate “solutions” may be evaluated using socio-economic criteria.

Given the power of GIScience for the exploration of data, it has become common to create systems that efficiently lay out alternative scenarios so that informed judgments can be made. This field has come to be called Spatial Decision Support Systems (Malczewski 1999). Public policy can be enhanced if various alternatives are clear not only to the policy makers but also to their constituents. These systems are designed to take complex spatially based problems and lay them out so that intelligent decisions can be taken.

GIScience and anthropology/archaeological research

To what extent did ancient roads facilitate exchange among prehistoric societies? The first task is to recreate the road network of the ancient society being studies. This might require the use of current remotely sensed data where road networks may still be observed but only from high altitude. Cost-path analysis, available in GIS, could be used to identify the possible links between settlements. It may be the result of such a study that trade was not a significant factor in early settlement patterns, but other factors such as religious rituals and political control were more important (Wescott et al. 2000).

Recent, ongoing studies of archaeological sites in the Mediterranean Sea and the Andes Mountains illustrate how GIS can be used to explore the legacy of the past at remote sites. The University of Haifa’s Maritime Civilization Department and the Geography Department’s GIS and Remote Sensing Lab are using a marine GIS to answer questions about maritime trade at Tel Shiqmona, located on the southern tip of Haifa Bay in Israel, during the period 538 B.C.–332 B.C. The marine GIS incorporates and integrates data on a number of variables, including coastal geomorphology, tides, currents, sedimentation, and sediment transport.

Since 1999, researchers from the University of California at Santa Barbara have been using GIS-based methods to record excavations at an archaeological site called Jiskairumoko in the Andes Mountains of Peru. Excavations and the recording of site data are typically extremely time consuming tasks, but the UCSB researchers are demonstrating that utilizing GIS for collecting and organizing data directly in the field permits the speedier excavation of a large area of the site and a highly accurate recording of buried archaeological deposits.

THE FUTURE OF GISCIENCE AND APPLIED GEOGRAPHY

One of the side effects of the introduction of GIS into mainstream geographic analyses and data exploration is that societies exposed to GIS are becoming more geographically literate. Many people now have a better sense of space, distance, direction, landforms, climate, political entities, and so on. This salubrious effect encourages more people to turn to geographic solutions to problems. The epidemiologist, economist, health care specialist, or political scientist will naturally think of spatial relations when dealing with georeferenced data, and, georeferenced data will more often be collected, manipulated, and presented. Specialists from several fields will be needed in order to fulfill the needs of the research. This will result in better communication among scientists and standardization in GIS notation and manipulation processes. Better techniques of data gathering, such as the use of computers in the field, more refined satellite images, and a host of new data manipulation technologies will only draw more attention to geographic solutions. In addition, more systems are becoming interoperable with regard to those produced in other parts of the world.

Perhaps the greatest fear in this otherwise positive view is that, as in so many of the new technologies, GIScience may be abused. Scenarios in an SDSS, for example, may be manipulated in such a way as to favor special interest outcomes. The issue of privacy becomes critical, especially when more detailed data are desired. Misinterpretation of results is always a problem. With GIScience, the issue of uncertainty in the data gathering instruments, data manipulation procedures and inappropriate tests on hypotheses are of considerable concern (Heuvelink et al. 1998, Hunsaker et al. 2001). For example, with finer data, the effect on results of spatial autocorrelation becomes magnified. Researchers need to be well informed of this phenomenon, otherwise biased or scientifically faulty results will obtain. There is always the danger that data gathered for one purpose may used unquestioningly for some other purpose. The problem of cost is also an issue. Most of the new technology requires large start up costs for equipment, software, and labor. This may result in having the rich and powerful control data gathering and data analysis techniques. Clearly, for the sake of good science, open systems and freely exchanged information are highly desirable goals.

ARTHUR GETIS
San Diego State University

REFERENCES

Anselin, L. (1988) *Spatial econometrics: methods and models*. Dordrecht: Kluwer.

Anselin, L. (1994) Local indicators of spatial association – LISA. *Geographical Analysis* 27: 93–115.

Anselin, L. and Rey, S. (1991) Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23: 112–31.

ArcNews ESRI, 380 New York Street, Redlands CA 92373–8100.

Bailey, T. C. and Gatrell, A. C. (1995) *Interactive Spatial Data Analysis*. Essex: Longman Scientific and Technical.

Bateman, I. J., Lovett, A. A., and Brainard, J. S. (2003) *Applied Environmental Economics: A GIS Approach to Cost-benefit Analysis*. Cambridge: Cambridge University Press.

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

Birkin, M. et al. (1996) *Intelligent GIS: Location Decisions and Strategic Planning*. New York: John Wiley & Sons.

Boots, B. and Getis, A. (1988) *Point Pattern Analysis*. Newbury Park, CA: Sage.

Briggs, D. J., Forer, P., Jarup, L., and Stern, R., eds. (2002*) GIS for Emergency Preparedness and Health Risk Reduction*. Dordrecht: Kluwer.

Clark, P. J. and Evans, F. C. (1954) Distances to nearest neighbor as a measure of spatial relationships in populations. *Science* 121: 397–8.

Cressie, N. (1991) *Statistics for Spatial Data*. Chichester: John Wiley.

Cromley, E. and McLafferty, S. (2002) *GIS and Public Health*. Guilford Press.

Diggle, P. J. (1983) *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.

Getis, A. (1984) Interaction modeling using second-order analysis. *Environment and Planning A* 16, 173–183.

Getis, A. (1995) Spatial filtering in a regression framework: experiments on regional inequality, government expenditures, and urban crime. In Anselin, L. and Florax, R. J. G. M. (eds.) *New Directions in Spatial Econometrics.* Berlin: Springer: 172–88.

Getis, A. and Ord, J. K. (1992) The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24: 189–206.

Getis, A. and Griffith, D. A. (2002) Comparative spatial filtering in regression analysis, *Geographical Analysis*, 34, 2, 130–140.

Goodchild, M. F. and Janelle, D. G. (2003) *Spatially Integrated Social Science: Examples in Best Practice*, Oxford: Oxford University Press.

Greene, R. W. (2002) *Confronting Catastrophe: A GIS Handbook*. Redlands: ESRI Press.

Griffith, D. A. (1988) *Advanced Spatial Statistics: Special Topics in the Exploration of Quantitative Spatial Data Series*. Dordrecht: Kluwer.

Griffith, D. A. (1996) Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *The Canadian Geographer* 40: 351–67.

Haining, R. (1990) *Spatial Dta Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.

Heuvelink, G., Goodchild, M. F. and Heuvelink, B. M. (1998) *Error Propagation In Environmental Modelling With GIS. London:* Taylor & Francis.

Hubert, L. J. (1979) Matching models in the analysis of cross-classifications. *Psychometrika* 44: 21–41.

Hunsaker, C. T., Goodchild, M. F., Friedl, M. A., and Case, T. J. (eds.) (2001) *Spatial Uncertainty in Ecology: Implications for Remote Sensing and Gis Applications.* Berlin: Springer-Verlag.

Jones, III, J. P. and Casetti, E. (1992) *Applications of the Expansion Method.* New York: Routledge.

Lang, L. (1999) *Transportation GIS*. Redlands: ESRI Press.

Longley, P., Goodchild, M. F., Maguire, D. J., and Rhind, D. W. (eds.) (1999) *Geographical Information Systems* (two volumes). New York: John Wiley.

Malczewski, J. (1999) *GIS and Multicriteria Decision Analysis.* New York. John Wiley.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209–20.

Okabe, A., Boots, B., and Sugihara, K. (1992) *Spatial Tesselations: Concepts and Applications of Voronoi Diagrams*. New York, John Wiley.

Ord, J. K. and Getis, A. (1993) Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis* 27: 286–306.

Paelinck, J. and Klaassen, L. (1979) *Spatial Econometrics*. Farnborough: Saxon House.

Ripley, B. D. (1981) *Spatial statistics*. New York, John Wiley.

Skidmore, A. and Prins, H. (2002) *Environmental Modelling With GIS and Remote Sensing*. 2nd ed., London: Taylor & Francis.

Wescott, K., and Brandon, R. J. (eds.) (2000) Practical Applications of GIS for Archaeologists: A Predictive Modeling Toolkit. London: Taylor & Francis.