SPATIAL ANALYSIS WITH GIS: PROBLEMS AND PROSPECTS

Michael F. Goodchild

National Center for Geographic Information and Analysis

University of California
Santa Barbara, California 93106

ABSTRACT

In principle, GIS provides an ideal platform for supporting a wide range of analyses using geographic data. In practice, the linkage of GIS with analysis is impeded by numerous factors, and has been more successful in some fields than in others. This paper looks at the problems of integrating GIS with analysis in general, and at the prospects for greater integration in the future.

The paper presents a broadly based classification of methods of spatial analysis. Efficient support of any class requires that the appropriate data model be recognized by the GIS. In some cases, methods of analysis are written for continuous space, without explicit discretization, and thus cannot be implemented intact. Many methods of spatial analysis require a data model that abstracts space to a simple matrix of interactions between objects; the paper discusses the implications of this class for the design of GIS, and argues for the development of simple 'hooks'. In other cases, the data model required to support analysis includes features such as time, or the vertical dimension, that are not commonly available in current GIS.

BACKGROUND: GIS AND SPATIAL DATA ANALYSIS

There seems to be widespread agreement in the GIS community on two simple propositions: that as a technology, GIS has the potential to support many different types of analysis; and that this potential has not yet been realized. This theme is reflected in Openshaw's oft-quoted comment:

> "Such systems are basically concerned with describing the Earth's surface rather than analysing it. Or if you prefer, traditional 19th-century geography reinvented and clothed in 20th-century digital technology." (Openshaw, 1987 p. 431)

It is also recurrent in the collection of papers on GIS edited by Worrall (1990).

The potential to support analysis is reflected in many discussions of GIS:

> "Geographic information systems evolved as a means of assembling and analyzing diverse spatial data." (Star and Estes, 1990 p. 14)

> "...the Geographic Information System...is as significant to spatial analysis as the inventions of the microscope and telescope were to science, the computer to economics, and the printing press to information dissemination." (Department of the Environment, 1987 p. 8)

But from a strictly pragmatic viewpoint, the reality of GIS today might be summed up in the following:

> A database containing a discrete representation of geographical reality in the form of static, two-dimensional geometric objects and associated attributes, with a functionality largely limited to primitive geometric operations to create new objects or to compute relationships between objects, and to simple query and summary descriptions.

GIS clearly needs stronger analysis and modeling capabilities if it is to meet its potential as a tool.

The case for Spatial Data Analysis (SDA) rests on the argument that explanation, understanding and insight can come from seeing data in their spatial context. There seem to be at least four separate arguments for this spatial perspective. First, space can provide a simple and useful indexing scheme. An archaeologist, for example, might record the locations of artifacts as they are unearthed at a dig simply in order to index them for later access. Geographic coordinates provide a kind of hashing code, on the assumption that two artifacts are unlikely to be unearthed at exactly the same location. A map provides a simple means of displaying the index, and finding artifacts given the human eye's extraordinary power to digest two-dimensional information. Spatial indexing is not likely to lead directly to insight, but it is a useful tool for handling large amounts of data.

Second, the spatial perspective allows easy access to information on the relative locations of objects and events, and proximity can indeed suggest insight. The Snow map showing the clustering of cholera victims in London during an outbreak of the disease in 1854 led directly to explanation, in the form of drinking water from a polluted well, and to an effective remedy for the outbreak (Gilbert, 1958). Although the explanation was immediately apparent from the map, it would have been virtually impossible to arrive at the same insight in any other way.

Third, a spatial perspective allows events of different types to be linked, in a process formalized in GIS as overlay. The fact that an event occurs in proximity to other events or objects can be very suggestive. For example, any environmental abnormality in the vicinity of a cluster of cancer cases is immediately suspect, because an individual is clearly more vulnerable to the local environment than to distant ones.

Finally, the distance between events or objects is often an important factor in interactions between them. In the physical sciences, distance can be a cause in itself, as in the inverse square laws of gravitation or electromagnetics. In the social sciences it is more likely to be a surrogate for information (we are more likely to know about nearby places than about distant ones) or for contact (we know more people locally) or time spent traveling (we prefer local shops to distant ones, all other things being equal) (Gatrell, 1983).

Spatial Data Analysis is a set of techniques devised to support a spatial perspective on data. To distinguish it from other forms of analysis, it might be defined as a set of techniques whose results are dependent on the locations of the objects or events being analyzed, requiring access to both the locations and the attributes of objects (Goodchild, 1987). Its techniques range from simple measures of the dispersion of a set of points to complex statistical tests of whether a set of points could have been generated by specific random processes (Ripley, 1981; Getis and Boots, 1978).

SDA provides a set of objective techniques to replace and augment subjective intuition. Unfortunately, while the spatial perspective can be very powerful as a source of insight, it can also be highly misleading. Ancient cultures found endless images in the random patterns in the night sky. More recently, there are many examples in the literature of false inferences drawn from apparent spatial patterns that later turn out to be no different from the outcomes of random processes. Haggett and Chorley (1969), for example, found that the average number of edges in the Brazilian administrative boundary network is very close to 6, and concluded that this supported the contentions of Christaller's Central Place Theory. But the bubbles in a polystyrene coffee cup also have very close to six edges on average, and analysis later showed that this is a necessary outcome of a theorem of Euler applying to any boundary network (Getis and Boots, 1978). Only this kind of objective analysis can determine whether visual cancer clusters are in fact abnormalities, or simply random events.

GIS needs SDA if it is to reach the potential implied by many of its definers and proponents, of a general-purpose tool for delivering a spatial perspective on data in a digital environment. SDA needs GIS if it is take advantage of the capabilities in GIS for data input, editing, display and mapping, and to be readily accessible to a broad user community.

PROGRESS TO DATE

If the arguments for linking GIS and SDA are so strong, why has so little progress been made to date? First, developments in the GIS industry largely reflect the demands of the GIS marketplace, which has been dominated for the past decade by applications in resource management, infrastructure and facilities management, and land information, where GIS tends to be used more for simple record-keeping and query than for analysis. Although a small minority of companies have stressed analysis in their development and marketing, SDA has had little impact on the GIS mainstream. SDA is also of greater interest in academic and scientific applications of GIS than in local government or the private sector, where GIS

budgets and expenditures tend to have been much higher.

Second, despite its promises, SDA remains a comparatively obscure field. There are few books or reviews, and there is no easy way of organizing or codifying SDA. Notable exceptions include the now classic collection edited by Berry and Marble (1968), Unwin (1981), Upton and Fingleton (1985) and the recent book by Haining (1990). There are no widely accessible courses in SDA, and there is some concern that the introduction of GIS into many university programs may in fact have diverted resources from existing courses in SDA (Heywood, 1990). In the long term, linkage with GIS may lead to greater awareness of SDA and greater availability of courses and texts, but in the short term there is a distinct shortage of knowledge, experience and training on the SDA side.

Third, many techniques of SDA were developed in the 1960s and 1970s when GIS was still in its infancy, and cartography a technology of pen and paper. Early efforts to implement SDA in a computational environment had to rely on source code programming, notably in Fortran (for examples see Baxter, 1976; MacDougall, 1976). Although the 1970s saw the emergence of integrated statistical packages like SAS and SPSS, with no explicit support for coordinates or spatial objects except for mapping and display, in practice these provided the most readily available basis for implementation of SDA until the recent interest in GIS. In addition, the statistical techniques on which much of SDA is based are in many cases explicitly non-spatial, making assumptions about the lack of spatial dependence which fly directly in the face of the spatial perspective. Spatial autocorrelation has often been treated as a problem to be removed (Odland, 1988), rather than as an inescapable property of almost all spatial data.

It is only in the past few years, with the development of GIS, that a realignment of SDA with other explicitly spatial technologies like cartography, GIS and remote sensing has finally begun. But much SDA remains strongly linked to the aspatial environment of statistical packages like MINITAB and SAS (Griffith, 1988).

Fourth, many techniques of SDA are complex and difficult, requiring a very different approach from the intuitive, synthetic view often promoted for GIS. A glance through the pages of a journal like Geographical Analysis is sufficient to strike despair into the hearts of most GIS enthusiasts. As an academic specialty with little immediate connection with the world of practical application, SDA might be accused of emphasizing mathematical sophistication at the expense of practicality. Simple, intuitive techniques for exploring data in a spatial context have often been ignored in the search for elegant formulations. Analysis, particularly when intuitive, is often associated with the inductive approach to quantitative geography that fell out of favor in the 1960s, giving way to modeling and deduction. The move to mathematical modeling has also been one form of response to the critique of positivism launched by the social theorists in the 1970s (Gregory, 1978). This chain of thought suggests that the key to integrating GIS and SDA may lie in an emphasis, at least initially, on the more intuitive, exploratory techniques in the SDA toolkit.

THE ROLE OF DATA MODELS

A GIS database captures real geographic variation in the form of a finite number of discrete, digital objects. Because geographical variation is fundamentally continuous and infinitely complex, this process of capturing reality must involve abstraction, generalization or approximation. The rules by which the objects and their relationships are defined is termed a data model (Tsichritzis and Lochovsky, 1977). The variety of data models used in GIS is one of its complications and at the same time one of its strengths.

Reviews of GIS data models have been published by Peuquet (1984) and Goodchild (1991). Data models take two broad forms, depending on whether reality is perceived as an empty space populated by objects, or as a set of layers or fields, each defining the spatial variation of one variable. In very broad terms, the former view is more relevant to analysis and modeling in the social sciences, where discrete entities are conceived as interacting over space, and the latter is more relevant to the environmental and physical sciences, but exceptions abound.

Objects are normally modeled as points (P), lines (L) or areas (A), after appropriate generalization of form, for example by representing a city as a zero-dimensional point. Fields are modeled in GIS in at least six ways:

- a raster of cells, each defining the average value of the field within the cell (e.g. a remote sensing scene) (R1F);
- a raster of regularly spaced point samples (e.g. a digital elevation model) (R2F);
- a set of non-overlapping, space-exhausting polygons, each defining a class (e.g. a soil or vegetation cover map) (AF);
- a set of irregularly spaced point samples (e.g. a weather map) (PF);
- a set of digitized isolines (e.g. a contour map) (LF);
- a set of non-overlapping, space-exhausting triangles, each assumed to approximate elevations within the triangle with a simple plane (the triangulated irregular network or TIN model) (TF).

Data models define how geographic variation is represented, but also determine the set of processes and analyses that can be carried out. For example, it would be appropriate to use a set of point samples representing a field of atmospheric temperature (the PF model above) to interpolate a contour map or create an oblique view, using the attributes of each point to determine the elevation of the interpolated surface. But it would be meaningless to perform the same operation on a set

of point objects (the P model) representing cities with attributes of population. Despite this, the two models may be stored identically in the GIS and the user may be unaware that a potentially meaningless operation is being performed.

Data models provide a logical and useful way of organizing the functionality of a GIS. They may also provide a framework for discussion of methods of SDA, since these are in principle extensions of basic GIS functionality. However many methods of SDA treat the issue of data modeling as a matter of implementation, rather than as an intrinsic property of the method of analysis. For example, suppose that a hydrological analysis requires the determination of ground slope, as an important factor in soil erosion. Slope is a well-defined property of any continuously differentiable mathematical surface. But it is not well-defined everywhere on the real landscape, which is characterized by frequent breaks of slope, and it is not defined independently of data model in any discrete representation of the land surface. In the TIN model, for example, it is well-defined and constant within triangles but indeterminate on their edges, and curvature is everywhere zero or indeterminate. In a contour model of the same real surface, slope must be inferred by some additional, as yet undefined process of interpolation.

In summary, an additional problem facing any effort to integrate GIS with SDA is that data modeling must be explicit in any use of GIS, but is often left undefined in SDA. Use of GIS forces the analyst or modeler to confront the issue of discretization directly.

DATA MODELS FOR SPATIAL DATA ANALYSIS

Although a wide range of data models are currently found in various GIS products, the range required to support a full array of spatial data analysis is much larger. Techniques of SDA can be arranged into several broad groupings depending on the underlying data model that is assumed in each one:

- points: techniques used to analyze an undifferentiated set of points, e.g. point pattern analysis (Getis and Boots, 1978);
- spatial objects with attributes: techniques that analyze an attribute matrix, and reduce space to a square matrix of spatial relationships between pairs of objects, e.g. measures of adjacency or proximity;
- networks of links and nodes: a range of techniques for analyzing networks in transportation and hydrology, based on attributes of network links and nodes;
- spatial interaction models: models of the interaction between pairs of objects, based on an analysis of the characteristics of origin objects, destination objects, and the spatial separation between them;
- raster techniques: methods of analysis based on the

representation of continuous layers as rasters of cells, and supported by the so-called raster GISs (a codification of this class has been developed by Tomlin, 1990).

Of these, the second and fourth require a matrix of relationships between objects that is missing in most currently supported GIS data models. Beyond these simple methods lie all of those models and techniques of analysis that require access to time or to the third spatial dimension. In other words, the current range of GIS products is far from adequate for supporting a full range of methods of spatial analysis.

THE NATURE OF A LINKAGE

One might define three different levels of linkage between GIS and SDA. Full integration would mean a common functionality accessible through a common interface, with associated conceptual structures. This seems unlikely to emerge given the nature of the GIS software industry and the unbounded nature of SDA. Close coupling would mean the recognition of common, high level structures in both systems so that information passed out of GIS and processed in SDA could be remerged without difficulty. For example, close coupling would require that the identities of objects be preserved when passed between GIS and SDA packages, so that if the order of objects changed, their identities would not be confused. To realize such close coupling and preservation of high level structures, the SDA system would have to know all of the data models in use in the GIS, which means in effect that it could not be a standard statistical package. Finally, loose coupling would mean that high level structures would be lost on transfer, and would thus have to be rebuilt on an ad hoc basis. This is the form of coupling that characterizes the relationship between many current GISs and other components of software federations, such as the statistical packages. Data must be transfered largely in the form of flat ASCII tables.

Practically, it seems that close coupling offers the most realistic alternative for improving on the current situation, which is characterized by loose coupling. For this to be achieved, however, there will have to be a much wider recognition of the role played by data models in GIS, and by the need to make discretization explicit in all aspects of spatial data analysis.

ACKNOWLEDGMENT

The National Center for Geographic Information and Analysis is supported by the National Science Foundation, Grant SES 88-10917.

REFERENCES

Baxter, R.S. 1976, Computer and Statistical Techniques for Planners, Methuen, London.

Berry, B.J.L. and Marble, D.F. 1968, Spatial Analysis: A Reader in Statistical Geography, Prentice-Hall, Englewood Cliffs, NJ.

Department of the Environment 1987, Handling Geographic Information. Report to the Secretary of State for the Environment of the Committee of Enquiry into the Handling of Geographic Information, HMSO, London.

Gatrell, A.C. 1983, Distance and Space: A Geographical Perspective, Oxford University Press, New York.

Getis, A. and Boots, B. 1978, Models of Spatial Processes, Cambridge University Press, London.

Gilbert, E.W. 1958, Pioneering maps of health and disease in England. Geographical Journal, Vol. 124, pp. 172-183.

Goodchild, M.F. 1987, Towards an enumeration and classification of GIS functions. Proceedings, IGIS 87: The Research Agenda, NASA, Washington, DC, Vol. II, pp. 67-77.

Goodchild, M.F. 1991, Geographical data modeling. Computers and Geosciences (in press).

Gregory, D. 1978, Ideology, Science and Human Geography, Hutchinson, London.

Griffith, D.A. 1988, Estimating spatial autoregressive model parameters with commercial statistical packages. Geographical Analysis, Vol. 20, pp. 176-186.

Haggett, P. and Chorley, R.J. 1969, Network Analysis in Geography, Edward Arnold, London.

Haining, R.P. 1990, Spatial Data Analysis in the Social and Environmental Sciences, Cambridge University Press, New York.

Heywood, I. 1990, Geographic information systems in the social sciences. Environment and Planning A, Vol. 22(7), pp. 849-854.

MacDougall, E.B. 1976, Computer Programming for Spatial Problems, Wiley, New York.

Odland, J. 1988, Spatial Autocorrelation, Sage, Newbury Park, CA.

Openshaw, S. 1987, Guest editorial: an automated geographical analysis system. Environment and Planning A, Vol. 19, pp. 431-436.

Peuquet, D.J. 1984, A conceptual framework and comparison of spatial data models. Cartographica, Vol. 21, pp. 66-113.

Ripley, B.D. 1981, Spatial Statistics, Wiley, New York.

Star, J.L. and Estes J.E. 1990, Geographic Information Systems: An Introduction, Prentice-Hall, Englewood Cliffs, NJ.

Tomlin, C.D. 1990, GIS and Cartographic Modeling, Prentice-Hall, Englewood Cliffs, NJ.

Tsichritzis, T.C. and Lochovsky, F.H. 1977, Data Base Management Systems, Academic Press, New York.

Unwin, D.J. 1981, Introductory Spatial Analysis, Methuen, London.

Upton, G.J.G. and Fingleton, B. 1985, Spatial Data Analysis by Example, Wiley, New York.

Worrall, L. 1990, Geographic Information Systems: Developments and Applications, Belhaven Press, London.