INT. J. GEOGRAPHICAL INFORMATION SYSTEMS, 1987, VOL. 1, NO. 4, 327–334

A spatial analytical perspective on geographical information systems

MICHAEL F. GOODCHILD

Department of Geography, University of Western Ontario,
London, Ontario, Canada N6A 5C2

Abstract. The field of geographical information systems (GIS) is reviewed from the viewpoint of spatial analysis which is the key component of the familiar four-part model of input, storage, analysis and output. Input is constrained by the limits of manual methods and problems of ambiguity in scanning. The potential for developments in output is seen to be limited to the query mode of GIS operation, and to depend on abandoning the cartographic model. Discussion of storage methods is organized around the raster versus vector debate and the need to represent two spatial dimensions in one. A taxonomy of GIS spatial analysis operations is presented together with a generic data model. Prospects for implementation are discussed and seen to depend on appropriate scales of organization in national and international academic research.

1. Introduction

It is often observed that the element which distinguishes geographical information systems (GIS) from other forms of spatial data handling activity, such as automated cartography and remote sensing, is an emphasis on analysis (Burrough 1986, Cowen 1987). Although there is clearly a very great variation in capability from one system to another, the ability to manipulate spatial data into different forms and to extract additional meaning from them is at the root of current interest in GIS technology. It seems appropriate, then, to review the present state of the field and assess progress to date from the perspective of spatial analysis. More specifically, this paper explores the extent to which current systems satisfy the ideal which, from this perspective, would consist of the ability to perform a comprehensive range of spatial analyses on all types of spatial data.

Because of the huge investments involved, the development of GIS tends to have been driven by application rather than by more abstract principles, and so the capabilities of current systems tend to reflect the needs of the commercial marketplace; in recent years defence and resource management seem to have been the major driving forces. The field appears to have paid relatively little attention to more idealistic visions of the significance of GIS. It is not surprising therefore that much academic research on GIS has focused on technical issues. To an outsider GIS research appears as a mass of relatively uncoordinated material with no core of theory or organizing principles. To some, this may be interpreted as indicative of the lack of significance of the field as a whole; to others, it is more a matter of the field's novelty, as theory and philosophy traditionally lag behind basic research activity.

This paper is organized in three major sections. The first reviews the field from the perspective of the traditional organizing principle of a four-part division (Tomlinson *et al.* 1976): input, output, storage and analysis. The next section focuses on analysis and presents a conceptual model for an ideal GIS from the perspective of spatial analysis. The final section evaluates the current state of the field against this model and discusses possible mechanisms for implementing it.

2. The four-component model

Many definitions of GIS make reference to the four basic functions of input, output, storage and analysis. From the perspective of spatial analysis the first three exist to support the last, which is seen as the real purpose of a system. The definition is useful in the sense that it more or less accurately reflects the investment of effort in system development; in fact, input and output have absorbed by far the greatest share of development costs in most current systems. This section focuses on the first three functions.

2.1. *Input*

Input functions continue to absorb a disproportionate share of operating costs in production systems (Goodchild 1987) and remain a major impediment to the adoption of GIS technology. The organizing principle distinguishing the various approaches which have been taken to input is the scale from manual to fully automated methods. Manual digitizing remains error-prone, despite the sophistication of much current software, and it appears unlikely that any significant improvement will be made over the present norm of about 60 polygons per hour in sustained production (Goodchild 1987). Complex rules are already used in many system to detect and rectify errors so it seems that techniques of artificial intelligence can offer relatively little further improvement.

Automatic digitizing is already cost-effective and must become more so as hardware becomes cheaper relative to manual labour and software becomes more sophisticated. There are already integrated scanning and vectorizing systems, capable of handling large, complex map documents, on the market for less than $100 000. However, the efficiency of scanning continues to be plagued by ambiguities in input documents. One possible solution lies in redesign; there is an analogy here to the redesign which occurred in many text founts to permit optical character recognition. The traditional map has evolved to satisfy the needs of the human eye and brain rather than the scanner/vectorizer. Simple changes in colours and use of such techniques as fluorescent inks and bar codes could make the topographic map far less ambiguous to a scanner, whilst remaining no less readily interpreted by the traditional user. In due course, with continued reduction in the cost of scanners driven by market growth, one might see redesigned paper maps replacing magnetic storage as back-up for large databases.

2.2. *Output*

The approaches taken to GIS output seem to fall more or less neatly into two modes, which will be referred to as product and query. The product mode is defined as an emphasis on hard copy in the form of maps, tables or lists, while in query mode the emphasis is on responses by the system to specific questions with optional hard copy. The importance of the distinction is not in its sharpness, since many systems clearly satisfy both modes to some extent, but in its implications for system design and for the adoption of GIS technology.

In product mode, it is assumed that the average user of spatial information is unwilling, for various reasons, to interact directly with the system. Instead, hard copy products which satisfy a broad range of possible query demands are generated periodically. It is assumed that the average user is able to specify what information should appear on these products and how often they should be generated (Goodchild and Rizzo 1987).

For example, a forest manager might schedule the generation of monthly reports on the maintenance status of all roads within the forest jurisdiction; decisions on maintenance would be made later as needed, by reference to the reports.

In query mode, the average user interacts directly with the system. In order to do so it is necessary for each user to maintain a high level of familiarity with the system and for the system to respond quickly to requests, whereas response time is relatively unimportant in product mode.

Although the distinction between query and product is not precise, there appears to be a growing division between them for reasons which are both technical and ~~operational. On the technical side, it is too difficult~~ for system designers to provide fast response times for very large databases over a broad range of functions; either functionality or response must be sacrificed. On the operational side, there appears to be a correlation between the narrowness of a user's demands on the system and his or her willingness to use it. If demands occur frequently and are all of similar nature, then the user is willing to maintain the level of knowledge of the system necessary for query mode, but if demands are relatively infrequent and broad in nature then the average user becomes frustrated with the system and returns to more traditional methods. The result is the adoption of query mode in agencies such as utilities and land records offices, requiring fast response to a limited set of questions, and product mode in resource management agencies where the mandate is more general. Vendors of GIS tend to specialize in one or the other because of the technical problems of satisfying both markets. Query mode systems have tended to employ data structures which use spatial as well as attribute-based indexing and present sophisticated user interfaces; product mode systems are often described as tool-boxes, with relatively unsophisticated interfaces but extensive functionality.

Because of its orientation toward conventional forms of decision-making, the product mode emphasizes the role of the map as the form of hard copy graphic output most acceptable to users. The query mode, on the other hand, has the potential to make use of novel techniques of graphic display. Although the map has evolved as a highly efficient means of communicating two-dimensional information, it reflects the technology which produces it, which is the pen, whether held by a human hand or directed by a digital computer. Real objects must be represented by pen strokes as points, lines of constant width or homogeneous areas delimited by precise boundaries. None of these models is particularly appropriate to the nature of the real objects being portrayed; because of the constraints of pen technology, maps are very inefficient at communicating time dependence, fuzziness and uncertainty, flows and any form of heterogeneity or continuous variation.

Product-oriented systems appear constrained by the need to generate hard copy maps, and even in some cases to emulate the techniques of traditional cartography in such areas as the placement of labels (Freeman and Ahn 1984). On the other hand, the query mode, with its electronic display, offers a radically new technology for communicating spatial information. High quality, ergonomically-acceptable, three-dimensional displays are now available and will make it possible to escape from the traditional two-dimensional constraints on display of three-dimensional topography. The third dimension can also be used, along with colour or density of shading, to generate bivariate maps. Another promising area lies in scene generation—the representation of spatial information by simulating its true appearance, rather than by symbolic cartographic objects. Recent developments in fractals, graphals and other stochastic process simulations of real objects (Fournier *et al.* 1982) make it possible to simulate complex real scenes, with potential applications in landscape architecture and studies of environmental change.

Digital spatial data contain uncertainty which can be attributed to a number of sources, such as inaccuracies and generalization in source documents and errors in digitizing (Walsh *et al.* 1987). Error was of little concern in traditional methods of map analysis because the imprecision of the analysis was almost always greater than the inaccuracy of the data. This situation is reversed in digital spatial analysis, and errors appear in such artifacts as sliver polygons. It has proved very difficult to deal with these errors because of the lack of any model of positional error for cartographic objects more complex than a single point. In part, this is because the objects themselves are abstractions, devised to satisfy the constraints of the input document. A boundary on a soil map is a cartographic representation of a zone of transition and, while it is relatively easy to model the transition of soil type across the boundary, it has proved much more difficult to find a suitable model of uncertainty in the boundary's position (Goodchild and Dubuc 1987). Yet electronic displays are forced to represent transitions as boundary lines. In a sense then, the error problem in GIS is artificial, a consequence of the continued use of cartographic objects for data input, storage and display.

2.3. *Storage*

The most basic organizing principle for the storage of spatial data is the distinction between raster and vector. Although many methods of digital representation have elements of both, it remains a fundamental point of variation between systems and data sets.

The persistence of the raster versus vector debate in the handling of spatial data seems to be due to its multidimensional nature which acts against a simple solution. First, the issue is one of resolution. Although the coordinates in a vector system must have a fixed level of resolution, determined by the machine's representation of numbers, it is often assumed that the precision of raster organization is fixed and predetermined, while that of vector organization is essentially unlimited. Database designers have often been unwilling to make the decisions on resolution which are required by raster systems, opting instead for the much higher precision of vector representation, which is almost certainly inconsistent with the accuracy of the data themselves. Unfortunately, the argument is not as straightforward as it appears. A vector system may be able to represent the apparent precision of pen-drawn maps, but its precision is often meaningless in relation to the data from which the map was derived and is therefore of no benefit. We lack any theory of raw data on which more objective and consistent choices could be made.

Secondly, the issue of raster versus vector is one of access, in the sense that a raster representation organizes the data according to spatial address, while a vector representation organizes them by object. Most systems now contain elements of both, in the interests of rapid access to the data according to a number of different kinds of query. This representation may take the form of a regular tiling system superimposed on a vector database, so that all data relevant to a particular tile are held together, or a more elaborate indexing of objects by location, using a bit-interleaved index. In general, product-oriented systems have adopted tiling while query-oriented systems have adopted dual spatial and object indexes in the interests of rapid response times.

Thirdly, the issue is one of data representation. A vector system tends to be chosen when the data are fully interpreted in the form of features, and it is desirable to link all components of a feature together for rapid analysis. Raster-based representation is more appropriate for uninterpreted data, such as digital elevation models (DEMs) and remotely-sensed images, where the data value in each pixel is a sample of some continuously varying field. If interpretation of imagery is to be followed by complex analysis, then it is often desirable to change to vector representation at some appropriate stage. For example, a road network might be detected from imagery in raster form, but an analysis of routing through the network (Lupien *et al.* 1987) would be much more efficient if the data were re-organized to an object orientation based on complete links.

The fourth dimension in the complex issue of raster versus vector is the question of algorithms. Some, such as the determination of area seen from a point or the optimal route between given origin and destination over a continuous cost surface, work efficiently on raster representations, and it is often assumed that polygon overlay is much more easily executed in raster than in vector representation. However, the case for the first example seems to be based more on the appropriateness of a raster for representing a continuous surface. In the second case of polygon overlay, the near-optimal vector algorithms now available can find all intersections between two sets of polygons in time proportional to the total number of polygons, which makes them highly competitive with raster overlay.

The final dimension is concerned with sampling; a vector data set represents a variable intensity of sampling whereas a raster implies a uniform intensity. One would therefore favour raster representations for examples such as remote sensing, where there is no possibility of varying sampling intensity in response to the complexity of the phenomenon, and vector representations in cases such as socio-economic reporting zones which must respond to the varying intensity of human settlement.

Recently, a consensus seems to have emerged on the raster versus vector debate. The type of data dictates the appropriateness of regular or irregular sampling, the desirability of object representation and the question of resolution. Both object and spatial indexing are required for rapid retrieval in response to queries. Finally, the algorithm issue is seen as largely irrelevant, given efficient algorithms. Rasters will be used for images and DEMs, but there should be a move to vector representation after interpretation or feature extraction.

An alternative way of looking at the raster versus vector debate is to interpret both as methods of reducing two-dimensional variation to one dimensional, either by sequential or random scan of the information. In this sense, the traditional row-by-row order of a raster is only one possible scan of a regular grid. It is clearly desirable to adopt an ordering which places neighbouring pixels close together in the data structure so that maximum advantage can be taken of runs in compressing the data and so that the data for a specified window can be retrieved in as few operations as possible. The Morton order, which scans in a sequence described by interleaving the bits of a binary representation of the (x, y) address of each cell, was devised to satisfy this objective. The Peano scan is more difficult to define but may have other advantages. Both are examples of a generalization in which a two-dimensional image is reduced to one by interleaving the digits of some representation of its (x, y) addresses.

3. Spatial analysis

If a database contains a digital representation of spatial data, then its data structure is, by definition, a model of spatial information. Thus we can interpret the research which has been undertaken in the past three decades on GIS data structures as an attempt to find an optimal model of spatial information. The criteria on which such a model is judged clearly involve the comprehensiveness of the model in embodying all types of spatial objects and the relationships between them, together with the efficiency with which the data can be analysed. The need to develop appropriate data structures can be seen as a powerful incentive to define a model of spatial information.

The data model which underlies such statistical packages as the Statistical Package for the Social Sciences (SPSS) and SAS (Statistical Analysis System) is very simple: objects occupy the rows and attributes and variables occupy the columns of a simple table. Nevertheless, the model is capable of supporting a full range of univariate and multivariate statistical methods. By analogy, a GIS data model must be capable of supporting a full range of spatial analysis.

In this section the elements of such a model for GIS are developed first. The model should not be interpreted as a data structure, as it does not have to satisfy the objectives of a system designer for speed and algorithmic efficiency. Instead, it represents the spatial analyst's view of the data in a GIS or the form of organization necessary to support a comprehensive range of forms of spatial analysis.

The model recognizes the three topologically defined primitive types of spatial objects: point, line and area. Continuous surfaces are mathematical abstractions which may be represented as elevation attributes at irregularly sampled points or as regular arrays of elevations so the raster pixel is included as a fourth object-type. Objects on the map are grouped into classes of different types according to theme. Each class has a defined set of attributes with meaning unique to the class. Finally, some attributes are constantly present; examples include polygon perimeter area and centroid location, and line length.

In addition to primitive objects, spatial analysis recognizes the need to allow a full range of relationships between pairs of objects of the same class, and between pairs of different classes. In all cases, the distance between the two objects is a constant attribute of these object-pairs but others include flow, number of migrants or transport costs. The possibility exists of relationships between objects taken three at a time but there appear to be very few recognized applications.

With this simple data model, six basic classes of spatial analysis can now be listed.

1. Operations requiring access only to the attributes of one class of objects. In this case, the model reduces to a simple table and it is likely that the analysis can be handled by a statistical package.
2. Operations requiring access to both attributes and locational information for a single class of objects; examples include calculating simple spatial descriptive statistics such as location of mean centre and dispersion.
3. Operations which create object-pairs from one or more classes of objects.
4. Operations which analyse attributes of object-pairs; examples include spatial autocorrelation indices and nearest neighbour analysis.
5. Operations requiring access to attributes and locational information for more than one class of objects or object-pairs. Spatial interaction modelling requires access to origin and destination objects and attributes of associated object-pairs.
6. Operations which create a new class of objects from an existing class, including generation of Thiessen polygons from points or buffer polygons around line segments.

Not all types of spatial analysis fit equally well within this model. More complex structures such as trees or paths must be represented as object or object-pair attributes rather than as abstract data types. The model is also not particularly convenient for system design, as it permits direct coding of 1:1 and *n*:1 but not *n*:*n* or 1:*n* relationships. However, it seems adequate as a spatial analyst's view, allowing a framework for organizing an otherwise amorphous mass of spatial analytical techniques.

If this model represents the spatial analyst's view of the ideal GIS, then no current commercial product comes close to matching it. Vector-based systems typically offer only a subset of the three object primitives and there are no systems currently on the market which recognize the object-pair and its importance to spatial analysis. Furthermore, the capabilities of most current systems for sophisticated analysis are very limited. Well-known algorithms for optimal spatial search, such as location allocation, have, until recently, been virtually unknown in the GIS field. Yet the potential is vast, both in increasing the capabilities of GIS and its areas of application, and in providing the supporting tools which will lead to much more widespread application of spatial analysis.

4. Implementation

Unfortunately, there is very little economic incentive to design and build a GIS which incorporates the spatial analytical model described above. As we noted at the outset, GIS is application-driven and, because each application is to some extent unique, system designers tend to have targetted their products at specific market sectors. It was also noted earlier that the nature of each application dictates the choice between product and query mode and the form of data representation. In this market-place, it is unlikely that any vendor would set out to develop an ideal GIS aimed at the common ground between diverse applications. Finally, the demand for spatial analysis is simply not strong enough to justify the risks of commercial development.

On the other hand, there are very strong academic and research incentives to develop a GIS for spatial analysis. First, it would provide an immensely valuable teaching tool. Offering both fast response and broad functionality, an educational system would not need to handle very large data sets, so that many of the practical difficulties mentioned above would not apply. Secondly, it would provide the organizing framework which appears to be missing in spatial analysis and in turn help to identify gaps in the present body of technique. Thirdly, it would be possible, in a system developed for academic purposes, to address many of the more abstract issues often ignored in commercial systems and to use it as a test bed for further research. For example, although the theory of regression analysis in the presence of spatial autocorrelation has been known for a long time, no attempt has yet been made to incorporate it within an available GIS and the inadequacy of the data model of SAS or SPSS ensures that it will not appear as an option within general statistical packages.

The key component of the system would be the input, output and storage functions needed to support analysis. With these in place, addition of different forms of analysis would be easy and there would be a strong incentive for researchers in spatial analysis to provide their results in the form of modules which could be exchanged within the research community. It seems that the lack of a supporting model, in the form of an inexpensive spatial database, is a major impediment to implementation of much of the excellent theoretical work in spatial analysis which has occurred over the past three decades.

In both the United States and the United Kingdom, there appears to be a growing recognition of the need for cooperation and centralization in GIS research to take advantage of the economies of scale which exist only at the national or international levels. It would be naive to argue that the research community is capable of the kinds of cooperation necessary to design and implement the necessary input, output and storage functions. On the other hand, a programme of cooperative development might be organized between a national centre and the commercial GIS sector. With this in place, the national centre could also coordinate the exchange of analysis modules.

5. Concluding remarks

The emphasis in this discussion has been on defining and reviewing the field of GIS from the perspective of spatial analysis which is often believed to provide its principal justification. In reality, most contemporary GIS place far more emphasis on efficient data input and retrieval than on sophisticated analysis. In part, this situation is a reflection of demand and is self-fulfilling; effective spatial analysis requires powerful analytical tools and will not become better known until these are available. In part, it is behavioural, since much of the current widespread enthusiasm for GIS is undoubtedly due to the attractiveness of their graphical output and their use of high technology.

There have been references recently to a need to close the gap between the technical sophistication of spatial analysis and the relative crudeness of its applications (Openshaw 1987). GIS has often been seen as having a useful role to play, but its ability to do so is currently limited by the inadequacy of GIS data structures to support a full range of analysis. It is argued in this paper that a national or international level of organization would be able to promote the kinds of changes necessary to bring about improvements and that progress would be rapid once an adequate data structure and associated input and output functions were available. Many of the technical difficulties of building such a system are obviated in the academic environment where it is less important to deal with very large data sets.

References

Burrough, P. A., 1986, *Principles of Geographic Information Systems for Land Resources Assessment*. Monographs on Soil and Resources Survey No. 12 (Oxford: Oxford University Press).

Cowen, D. J., 1987, Computer mapping vs. GIS: implications for applied geography. *Papers and Proceedings of Applied Geography Conferences*, **10**, 43.

Fournier, A., Fussell, D., and Carpenter, L., 1982, Computer rendering of stochastic models. *Communications, Association for Computing Machinery*, **25**, 371.

Freeman, H., and Ahn, J., 1984, AUTONAP—an expert system for automatic map name placement. *Proceedings of the First International Symposium on Spatial Data Handling held in Zurich on 20–24 August 1984* (Zurich: Department of Geography, University of Zurich-Irchel), p. 554.

Goodchild, M. F., 1987, Application of a GIS benchmarking and workload estimation model. *Papers and Proceedings of Applied Geography Conferences*, **10**, 1.

Goodchild, M. F., and Dubuc, O., 1987, A model of error for choropleth maps, with applications to geographic information systems. *Proceedings of AutoCarto 8 held in Baltimore, Maryland on 29 March–3 April 1987* (Falls Church, Virginia: American Society Photogrammetry and Remote Sensing, American Congress of Surveying and Mapping), p. 165.

Goodchild, M. F., and Rizzo, B. R., 1987, Performance evaluation and workload estimation for geographic information systems. *International Journal of Geographical Information Systems*, **1**, 67.

Lupien, A. E., Moreland, W. H., and Dangermond, J., 1987, Network analysis in geographic information systems. *Photogrammetric Engineering and Remote Sensing*, **53**, 1411.

Openshaw, S., 1987, Guest editorial: an automated geographical analysis system. *Environment and Planning* A, **19**, 431.

Tomlinson, R. F., Calkins, H. W., and Marble, D. F., 1976, *Computer Handling of Geographical Data* (Paris: UNESCO Press).

Walsh, S. J., Lightfoot, D. R., and Butler, D. R., 1987, Recognition and assessment of error in geographic information systems. *Photogrammetric Engineering and Remote Sensing*, **53**, 1423.