GIS, Spatial Representation, and Statistical Mapping

*Goodchild, Michael F.

National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA. Phone: 805 893 8049. FAX: 805 893 7095. Email: good@ncgia.ucsb.edu.

Abstract

Statistical maps can be created from spatial databases containing attributes of point and area primitives. The simple GIS functions of point-in-polygon and polygon overlay can be used to aggregate point records, and to transfer attributes from one set of polygons to another. But current spatial database designs include capabilities for advanced data modeling. These include dynamic segmentation of linear networks, attributes of object pairs, hierarchical constructs such as parent-child relationships, probabilistic and fuzzy representation, and representation of temporal change. Some are handled within the georelational model, while others take advantage of newer concepts such as object orientation. The presentation reviews these concepts of advanced spatial data modeling, and examines their potential usefulness to the subject matter of the conference.

Introduction

In recent years, the field of geographic data modeling has emerged as a distinct and important subspecialty of geographic information science, and is now recognized as one of the most fundamental of the research issues raised by the development and exploitation of geographic information systems. In essence, geographic data modeling concerns the techniques used to create discrete digital representations of geographically varying phenomena within spatial databases and geographic information systems. Because real geographic variation is almost infinitely complex, decisions must be made in collecting, characterizing, and storing geographic information that will ultimately affect the types of analysis that can be performed on the information, and the success of scientific investigations into geographically distributed processes. Geographic data modeling can be seen as a subfield of data modeling in general, and also as a subset of the broader issues of geographic representation--the systematic study of the methods people use to create representations and understandings of the world around them.

At this point, it is perhaps worth stressing the distinction being made here between geographic data modeling, and the modeling of processes operating on the geographic landscape. In the first case the emphasis is on representation, and the choices that must be made to create a discrete digital version of some real geographic form or pattern. Process modeling, on the other hand, is concerned with the mathematical or numerical representation of processes rather than forms, and with prediction and understanding rather than with representation per se. Thus data modeling is a preliminary to science, although the representational choices made will undoubtedly influence the way scientific investigation is ultimately carried out.

Peuquet published a comparatively early and much cited review of geographic data modeling (Peuquet, 1984). Goodchild (1992) provides a more recent review, and the increasing importance of the field is reflected in the chapters devoted to this topic in the review of GIS by Maguire, Goodchild, and Rhind (1991). Notable among more specialized reviews are the discussions of modeling the temporal component in spatial databases by Langran (1992) and Peuquet (1994), and of three dimensional spatial information by Raper (1989) and Turner (1992). Smith et al. (1994) demonstrate the importance of geographic data modeling in environmental science, giving examples of the role data models can play in the formulation and use of mathematical models of environmental systems and processes.

Data modeling is a complex and often subtle process that extends from the design of surveys and instruments for capturing raw data through the various manipulations that occur before the data are finally ready to be analyzed or displayed. For example, in the case of the Census of Population, the decision to represent continuous change through time as a series of snapshots taken every ten years was made decades ago; a decision to represent spatially continuous change in average personal income by calculating one average per county might be made much later and specifically for the purposes of preparing a map display from a database of individual records; and the decision to represent the smooth curves of a county boundary along a river as a set of connected straight segments would have been made when the county boundary was digitized. Each of these decisions ultimately affect how the world is represented in the digital database, and constrain the kinds of analysis that can be performed, but they are made at different times, by different people, and perhaps with different objectives in mind.

The purpose of this paper is to address the geographic data modeling issue in the context of this conference, with particular emphasis on environmental health and the statistical mapping of health data. After a review of the use of traditional geographic data models, the paper outlines the principal findings of recent research on advanced data modeling, and their potential applications in this area. While many of these have yet to appear across the full range of available software products, their eventual availability is likely to have a substantial impact, in changing the ways in which health analysts and epidemiologists collect and represent geographic data, and structure models of processes.

Traditional GIS Data Models

Traditionally, GIS data models have been divided into two major classes: *raster* data models, which represent spatial variation by assigning values to cells in a fixed rectangular array; and *vector* data models, which represent spatial variation through irregularly distributed points, lines, and areas. Lines and area boundaries are often conceived as smoothly varying, but the vector data models represent both as connected sequences of straight segments (*polylines* and *polygons* respectively), although some systems also support smooth representations known as *splines*, and also connected arcs of circles. In a raster model, each cell in an array is normally restricted to a single value, so multiple *layers* of coincident cells must be used to represent the spatial variation of many variables. For example, a layer might represent the spatial variation of average income, calculated by census tract, or the pH of drinking water in a layer constructed by *interpolating* between a number of measurement locations.

The meaning of *layer* for vector systems is more complex. For some systems, a layer is simply a convenient grouping of one or more classes of points, lines, or areas; a layer might include all streets, for example, or all county boundaries. In such systems organization of the data into layers facilitates display, allowing the user to turn layers on or off, and assign them different colors or other types of symbolization. In other systems, the points, lines, or areas forming a layer must be *planar enforced*. In the case of a layer of areas, this means that areas cannot overlap and must fill the project area, so that every location within the project area lies in exactly one area. This is true in general of many classes of statistical reporting zones, such as census tracts or counties.

Because of this confusion, there have been several attempts to provide an overarching conceptual framework that can accommodate the various implementations found among available GIS and spatial database products. Goodchild (1992) argues that two views of spatial variation underlie the various options. One is *afield, defined as* the view *that a* variable can be given a single, well-defined value at every location within the project area. An example of a field variable is *elevation*. Many systems of statistical reporting are defined as assigning every location to a single zone--for example, every point in the US lies in exactly one county (if we ignore the possibility of lying exactly on a boundary). Counties thus satisfy the definition of a field, since they assign every location to a single value, measured in this case on a nominal scale as an identifying name.

While elevation and county satisfy the definition of fields, other spatial variables raise problems. It is not immediately obvious that population density or average income, for example, can be conceived as fields, although they are frequently displayed as continuous surfaces with a single value of the variable evident at every location in the project area. Population density can be defined for a county by dividing population by area, but this definition breaks down as the area of the reporting zone shrinks, and becomes meaningless for areas below perhaps 200m in diameter. But this problem can be avoided in two readily available ways. First, we could define population density as a field by specifying an explicit search radius--for example, value at a point could be defined as the ratio of the number of people found within 100m of the point divided by the area of a circle of radius 100m. Second, we could make use of techniques of density estimation, which replace discrete point locations with a spatially continuous *kernel function* (Silverman, 1986. Within the geographical literature density estimation is analogous to the calculation of potential; see for example Stewart, 1947).

For spatial variation conceived or characterized as a collection of fields, spatial databases currently provide six distinct methods of discrete, digital representation. The six field data models are:

- a regular rectangular array of cells, with the average value of the field recorded in each *(raster* model);
- a regular rectangular array of sample points *(grid* model);
- an irregularly spaced set of sample points *(point* model);

- a polyline representation of a set of isolines *(contour* model;
- a set of planar enforced irregular polygons, with associated average values for each *(polygon* model);
- a set of planar enforced irregular triangles, variation within each triangle being assumed to be planar *(triangulated irregular network* or *TIN* model).

Since the average or mean value cannot be defined for nominal-scaled data, it is normally replaced with the modal or commonest value, although in many cases the nominal value will be true for the entire cell or polygon (e.g., county name or census tract ID). The names suggested above are conventional, although many equivalent terms exist.

Of the six field representations, the polygon model is by far the most commonly used method for applications in epidemiology and environmental health. Statistics are often released on the basis of reporting zones with irregular boundaries, and increasingly such boundaries are already available to analysts in digital form. It is possible, for example, to obtain digital representations of ZIP boundaries, counties, states, and census tracts. GIS technology makes it easy to aggregate individual point records to ZIP boundaries using mailing address, and thus to estimate the value of field variables such as standardized mortality ratios.

On the other hand, the polygon model presents certain disadvantages. Many types of reporting zones display wide variation in area, and population may be far from evenly distributed within them. The case of San Bernardino County, the nation's largest county by area, is often cited since it includes vast areas of unpopulated desert in addition to heavily populated urban areas. The uniform cell size of a raster data model gives the potential to avoid such problems, provided estimates can be made reliably by cell. While there have been cases of data collection by rectangular cell (the 1981 Census of the United Kingdom provided data by 1km cell), a number of methods have also been devised for intelligently transforming data collected for irregular polygons to a regular grid. Tobler's pycnophylactic method (Tobler, 1979) is one example, and more recently Martin and Bracken (Martin and Bracken, 1991; Bracken, 1993) have shown the value of such techniques for exploring small area census statistics. An additional advantage of such raster representations is that they can be used to reaggregate statistics to a basis of new polygonal reporting zones (Goodchild, Anselin, and Deichmann, 1993).

Not all geographic phenomena are conveniently represented as fields, and the requirement that a field have a single value everywhere in the plane is often overly restrictive. Thus in addition to fields, GIS support the representation of phenomena as discrete objects--points, lines, or areas--which may overlap, and need not fill the plane. These *entity* models reflect an entirely different view of the world as composed of discrete objects with internally homogeneous characteristics, and one that is perhaps much closer to the ways people learn about the world and navigate within it (Mark and Frank, 1991). The concept of *neighborhood*, for example, is much better accommodated within the entity view of the world, since boundaries of neighborhoods may overlap, and are almost certainly specific to context. One can always create fields from such entities--for example, a single neighborhood can be represented as a field variable with two values, inside and outside, and many people's views of a given neighborhood can be represented as a variable equal to the proportion of respondents who indicated that a given location was inside. But while such representations may be worth creating within the constraints of any one spatial database architecture, they are generally inefficient because of the high level of redundant information.

Entity models are common among GIS architectures that have been designed for facilities management applications, such as are found in highway departments, city engineering departments, or utility companies, where there are many examples of discrete entities, and almost none of spatially continuous fields. Some commercial GIS offer no capabilities at all for storing or manipulating fields.

In environmental health applications, discrete entities are encountered in the form of the base map information needed to visually orient map displays, such as the locations of streets or cities. Point occurrences of disease are also conceived as point entities, although they can be used to compute fields of density as discussed earlier. While a point occurrence might be stored identically to a point sample from a continuous field in a spatial database, the operations that make sense for it are entirely different. Given a series of point weather observations, for example, one might want to interpolate a continuous surface, but this operation makes no sense whatever for a series of point occurrences of disease. On the other hand it is reasonable to conceive of a map of point occurrences as the outcome of a Poisson process parametrized by a field density variable, and to use point occurrences to try to estimate the density field; or to aggregate point occurrences based on a field of reporting zone identifiers.

The Georelational Model

The basis for any digital representation of geographic phenomena is a collection of points, lines, and areas, though the rules governing collections of these primitive objects depend on whether the entity or field view is being implemented. In addition to primitive objects, any successful approach to geographic data modeling must allow for the representation of relationships between objects. Such relationships fall conveniently into three categories: those necessary to the successful representation of the objects themselves; those describing relationships evident from the objects' geometry; and those describing relationships of a more functional nature. The first category will not always be apparent to the user, because its examples may have been hidden by the system designer, but it includes the relationships between adjacent areas in a planar enforced layer of such areas, and the relationships between shared vertex points that define a TIN. The second may be the most important category for environmental health applications. It includes the *lies in* relationship that assigns a point to its containing reporting zone, and the *overlaps* relationship that indicates whether the areas of two reporting zones intersect. Functions exist in most GIS to compute such relationships from the geometry of the primitive objects: The *lies in* relationship is computed with the *point-in-polygon* function, and *overlaps* with *polygon overlay*.

The *georelational model* emerged in the late 1970s as a specific implementation of the more general relational database model that offered specific advantages for storing and manipulating spatial relationships. It is now the dominant architecture for GIS, though its dominance may be threatened in the next few years by more recent alternatives. Spatial objects are organized into classes sharing common characteristics, such as a class of county polygons, or a class of point occurrences of disease. Each class is conceptualized as a table, with the member objects forming the rows and the various attributes of each object as the columns. This tabular model of data should be familiar from its use in the statistical packages and spreadsheets. In the georelational model, a relationship between two objects is represented as an additional attribute of one or sometimes both of them. For example, the *lies in* relationship would be represented as an additional column in the table of points, containing the identification number of the containing polygon, and thus *pointing* to its record in the relevant reporting zone table. There will be exactly one pointer because a point must lie in exactly one polygon, whereas the reverse, of pointing from polygons to points, would not be as simple. Simple operations on tables, such as counting points by polygon, are now easily carried out.

Besides relationships that can be determined from geometry, the georelational model also provides support for other kinds of relationships, such as those having to do with the processes occurring on the landscape and responsible for the patterns we see. One commonly implemented relationship in transportation applications of spatial databases is connectivity, allowing one link in the transport network to *point to* another link. Although these can often be inferred from geometry, the complications of overpasses, underpasses, and illegal turns force the builders of such applications to treat each possible connection explicitly.

Interactions

Although pointers allow relationships to be defined between different types of objects, or between objects belonging to a single class, they provide no support for the qualification of relationships. For example, one might want to store the fact that a particular connection between streets could only be made at certain hours of the day, or by certain types of vehicles. In an example application perhaps more relevant to environmental health and epidemiology, one might want to represent the amount of interaction, social or physical, between two different spatial objects, such as the number of migrants recorded between two counties, or the amount of groundwater flow between two wells. The georelational model allows us to do this by establishing a new table, or *relation*, with one entry per interaction. The columns of the table would include a pointer to the origin object, a pointer to the destination object, and various measures of interaction such as flow of people, distance, or travel time.

Unfortunately, at this point in time no available GIS provides comprehensive support for the representation of interaction. While relations such as that defined above exist in isolated instances, the necessary functions for basic manipulation of interaction data, such as display, or modeling of flows using standard spatial interaction models, are not provided by the majority of GIS vendors, and interaction modeling is only just beginning to appear as a supported function. Hopefully such functions will be much more common in the future.

The next section of the paper reviews recent work in advanced data modeling, and its potential applications to environmental health and epidemiology.

Advanced Geographic Data Modeling

Recent research has yielded a rich set of advanced techniques for geographic data modeling, particularly in areas such as time, linear networks, and hierarchies. The subsections of this part of the paper review each of the major contributions, and outline ways in which they can be useful to applications in epidemiology and environmental health.

Hierarchical structures

The traditional emphasis in geographic data modeling has been on the representation of a single level of generalization. Tools for automatic generalization, or the automatic production of one scale or spatial resolution of mapping from another, larger scale, have proven highly elusive (Buttenfield and McMaster, 1991), because of the difficulty of emulating in a machine the complex and sophisticated process used by the cartographer. Recently, however, significant advances have been made in the digital representation of hierarchy, or the logical linkage between data at different levels of spatial resolution.

One of the best known of these is the concept of the quadtree (Samet, 1990). Rather than model a single level of spatial resolution, a quadtree systematically decomposes a map into scales of variation, starting with the coarsest and proceeding to the most detailed. The basic quadtree concept has been implemented both as a method for representing fields, and as an index to the spatial objects in an entity database. In the latter case, the largest objects are associated with the coarsest spatial resolution. Advanced versions of this concept include more adaptable tree structures (von Oosterom, 1993) and wavelet decompositions (Chui, 1992).

Complex objects are another mechanism for modeling relationships between information at different spatial scales that is now widely supported among available GIS, particularly for facility management applications. Here, a feature represented by a single spatial object at one level of resolution may be linked to a set of more primitive features at another, more detailed level. This situation is commonly found in the systems of reporting zones used for health and socioeconomic data, where a state may be represented by a single polygon at one level, but as a set of polygons each representing a county at another level, and may perhaps decompose even further at still finer resolutions. Nested hierarchies such as this are common in geographical data, so it is surprising perhaps that so little support has been provided for them in traditional GIS designs. Support should include the ability to define relationships between objects at different scales, and simple functions such as aggregation and disaggregation. Recently, new techniques of exploratory spatial data analysis have allowed more than one level in a hierarchy to be viewed and analyzed simultaneously (Fotheringham and Rogerson, 1994), leading to useful insights into the effects of scale on the analysis of geographic data.

Although it is possible to model hierarchical relationships within a relational database, lack of supporting methods of manipulation and explicit techniques for representation have created the sense of a conflict between the georelational model as normally implemented in GIS, and the need to model hierarchical relationships. In the case of GDS, a GIS widely adopted within the facilities management industry, the central database architecture is hierarchical rather than georelational, indicating the importance attached to hierarchical spatial relationships in this application area, although links also exist to standard relational databases. Hopefully, new GIS products appearing on the market in the next few years will make the modeling of hierarchical relationships much easier.

Dynamic segmentation

Road and river networks have provided a rich set of applications for GIS. In traditional geographic data modeling, networks are represented as connected sets of lines. Each line object points to other connected line objects, or to the nodes (a special class of points) at each end; in the latter case, the nodes in turn point to links. Link/node modeling of transport networks began in the 1960s with the Census Bureau's DIME project, and is currently generating widespread interest in the context of Intelligent Vehicle-Highway Systems (IVHS). Information about the distribution of phenomena along the network can be represented in the form of link and node attributes. GIS applications include vehicle routing and scheduling, and site selection for facilities on the network.

Although successful, the link/node model suffers from one major disadvantage: in reality, attributes are often not constant along links, and change in attributes is not confined to nodes. Rather than introducing false nodes at every change of attributes, the *dynamic segmentation* model allows attributes to be assigned to points anywhere on the network, or to sections of links between points that can span intervening nodes. This advance on the traditional GIS data model has proven very useful in network applications wherever it is necessary to describe complex geographical variation over the geometry of a network.

Fuzzy representations

In the traditional GIS data model all variation must be expressed through some combination of discrete, well-defined objects with precise *footprints* on the surface of the Earth. In a vector system, precision is determined by the arithmetic precision of the computer and its programming language, and can often be as high as one part in 1014. In raster-based systems positional precision is determined by cell size and is thus likely to be much coarser, but it is rare to find rasters being used to represent positions of spatial objects in environmental health or epidemiological applications.

Despite the precision of vector-based systems, the positional accuracy of spatial objects is likely to be much less. Even the best surveying instruments achieve accuracies of only 1 part in 106, and map digitizing can rarely exceed 1 part in 104. Moreover, the definitions of many spatial objects fail to support even these levels of accuracy. A map of soils or land cover, for example, may include boundary lines between classes whose positions are definable to no better than 1 part in 100 (such as a boundary positioned to the nearest 100m on a map covering an area 10km wide). For this reason, there has been much interest recently in the modeling of uncertainty in spatial databases, and in studies of error and accuracy (Goodchild and Gopal, 1989). Fuzzy logic has been widely suggested as a framework for such models.

Such extremes of positional uncertainty may seem unlikely in epidemiology, although they may well occur in the context of certain issues in environmental health. On the other hand, such applications may require the merger of data sets with widely differing lineages and thus positional accuracies. To take one example, many boundaries of reporting zones run along major streets. To assign the resident of a house on the street to the correct reporting zone can require a positional accuracy of perhaps 20m in both the data set containing the digitized boundary polygons, and the data set containing the point records of residents. Such accuracies are common at mapping scales of 1:40,000 and larger, but are not achieved by poorly controlled street maps and other readily available products.

Although traditional GIS data models made no accommodation for uncertainty, requiring the assumption that all object footprints were *crisp*, there has been much recent interest in explicit support for uncertainty, documentation of information on accuracy, and programs of accuracy assessment. Fear of legal liability is one of the major motivations behind this new interest. The recently adopted Federal spatial data format (Federal Information Processing Standard 173; Morrison, 1992) and metadata (data documentation) standards both contain guidelines for reporting and accessing information on quality. Other recent research has focused on techniques for communicating quality information to the user, through novel methods of visual display (Beard, Buttenfield, and Clapham, 1991).

Other recent research effort has gone into methods of *error propagation*, which allow the effects of uncertainty in input data to be carried automatically into estimates of the uncertainty of the products of analysis. Such methods may be useful in estimating the uncertainties associated with spatially-based policies in the area of environmental health, and in dealing effectively with risk. Epidemiologists are likely to have their own approaches to uncertainty, based on statistical models, and it may be useful to explore their incorporation as GIS functions.

Temporal change

Raw data collection in the field is costly, and traditional systems of data collection thus tend to emphasize those aspects of real systems that remain relatively constant through time, and to characterize change through the use of periodic *time slices*. This approach is reflected in the organization of the decennial census, and also in such systems as Landsat, which takes a view of the surface of the Earth from space every 19 days. Maps also emphasize the comparatively static aspects of the Earth's surface, and these traditions have been preserved in the traditional GIS data models.

More recently, there has been substantial interest in introducing the temporal dimension to spatial databases, and in investigating whether traditional methods of data collection might not be usefully modified in response to the widespread use of digital data processing. Is it possible to rethink the timeslice approach to data collection in the age of GIS?

Langran (1992) and Peuquet (1994) provide recent reviews of the state of temporal data modeling in GIS. One commonly encountered problem is the instability of reporting zone boundaries through time, which creates problems for any kind of longitudinal data analysis. Even though longitudinal stability is a stated objective of the Census Bureau's reporting zones, changes nevertheless occur in boundaries due to urban development and population shifts, and they are much more common in other reporting zones such as ZIPs and postcodes (Raper, Rhind, and Shepherd, 1992). The postcode-based approach to the problem adopted in the UK has been to maintain the stability of the smallest geographic units (smaller than the typical city block), while allowing limited reaggregation in response to development. Another possibility is to maintain a master file of all boundary segments that ever existed, and to incorporate a GIS function that can rebuild from them the exact geometry of reporting zones that was valid at any specific time. This approach could be used, for example, to build the county boundary network of the US at any time since Independence.

Another approach to spatiotemporal data modeling is illustrated by the work of Goodchild, Klinkenberg, and Janelle (1993) in representing the daily behavior patterns of a sample of urban residents. Rather than time slices, the raw data consisted of the known locations of each individual in the sample; during periods of movement only the endpoints are known. This data model can be visualized as a three-dimensional space, with the two spatial dimensions horizontal and time vertical. In this space, each individual's trajectory forms a line having a single spatial location at any point in time. Simple processing functions can be used to interpolate a map of individuals at any time, or to aggregate individuals by spatial reporting zones for analysis.

In other applications, it may be necessary to represent the space-time continuum as a three-dimensional or even four-dimensional raster, in order to support models of space-time processes. Such approaches are likely to be expensive both in data acquisition and in processing.

Conclusions

The focus of this paper has been on the nature of geographic data modeling in two contexts. Traditional GIS data models have emphasized the representation of the contents of maps, and thus have modeled real geographic phenomena as layers of fixed, crisply defined points, lines, and areas. The advanced data models that have emerged in the past decade, largely in response to the inadequacies of the earlier generation, have dealt explicitly with time, fuzziness, hierarchical relationships across scales, interactions, and complex patterns on networks. These new ideas have only just begun to influence practice, and their full influence is yet to be seen. The GIS field is conservative in such matters, and it can be difficult for products offering new ideas in data modeling to break into a market dominated by a few large vendors. The paper has also made suggestions about areas where the newer more advanced data models are likely to influence practice in environmental health and epidemiology.

There has not been space in this paper to discuss several important trends in advanced data modeling. The paper has followed traditional practice in separating the issues of representation from the functions to be performed on the data, and this tradition is now under regular and strong attack from supporters of *object orientation,* who argue that by *encapsulating* functions with data it will be possible to escape many of the limitations currently imposed by choice of data model. For example, if a set of irregularly spaced point observations of air temperature could be packaged with a function for spatial interpolation, then it would be possible for the user to ignore the particular representation, and to think of the data simply as a field. This kind of approach holds the promise of much greater ease in transferring data from one system to another, and of *interoperability* between systems.

With a steadily increasing number of options for geographic data modeling available, it seems likely that *niche* implementations of GIS will appear, offering the subset of data modeling options found to be most useful in some specific application area. Dynamic segmentation is more likely to be supported by vendors competing in the market for transportation applications, while clearly of less interest for resource management. Epidemiology seems to be a sufficiently focused area of application to form such a niche, though it is far from clear at this point what particular advanced data modeling options will prove useful--hierarchical structures among reporting zones would seem an obvious candidate. On the other hand environmental health seems a much wider field, requiring support for temporal and perhaps even three-dimensional representations. Experience in other areas suggests that it will be some time before a strong consensus emerges on this question.

Acknowledgments

The National Center for Geographic Information and Analysis is supported by the National Science Foundation, grant SBR-8810917.

References

Beard, M.K., B.P. Buttenfield, and S.B. Clapham, 1991. *NCGu Research Initiative 7: Scientific Report for the Specialist Meeting.* Santa Barbara, CA: National Center for Geographic Information and Analysis.

Bracken, I., 1993. An extensive surface model database for population-related information concept and application. *Environment and Planning B* 20(1): 13-27.

Buttenfield, B.P., and R.B. McMaster, editors, 1991. *Map Generalization: Making Rules for Knowledge Representation.* London: Longman Scientific and Technical.

Chui, C.K., 1992. *An Introduction to Wavelets.* Boston: Academic Press.

Fotheringham, A.S., and P.A. Rogerson, 1994. *Spatial Analysis and GIS.* London: Taylor and Francis.

Goodchild, M.F., 1992. Geographical data modeling. *Computers and Geosciences* 18(4): 401408.

Goodchild, M.F., L. Anselin, and U. Deichmann, 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25: 383-397.

Goodchild, M.F., and S. Gopal, 1989. *Accuracy of Spatial Databases.* London: Taylor and Francis.

Goodchild, M.F., B. Klinkenberg, and D.J. Janelle, 1993. A factorial model of aggregate spatio-temporal behavior - application to the diurnal cycle. *Geographical Analysis* 25(4): 277-294.

Langran, G., 1992. *Time in Geographic Information Systems.* London: Taylor and Francis.

Maguire, D.J., M.F. Goodchild and D.W. Rhind, editors, 1991. *Geographical Information Systems: Principles and Applications.* London: Longman Scientific and Technical.

Mark, D.M., and A.U. Frank, editors, 1991. *Cognitive and Linguistic Aspects of Geographic Space.* Dordrecht: Kluwer.

Martin, D., and I. Bracken, 1991. Techniques for modelling population-related raster databases. *Environment and Planning A* 23(7): 1069-1075.

Morrison, J.L., 1992. Implementing the spatial data transfer standard - introduction. *Cartography and Geographic Information Systems* 19(5): 277.

Peuquet, D.J., 1984. A conceptual framework and comparison of spatial data models. *Cartographica* 21(4): 66-113.

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

Raper, J.R., 1989. *Three Dimensional Applications of Geographical Information Systems.* London: Taylor and Francis.

Raper, J.F., D.W. Rhind, and J.W. Shepherd, 1992. *Postcodes: The New Geography.* London: Longman Scientific and Technical.

Samet, H., 1990. *The Design and Analysis of Spatial Data Structures.* Reading, MA: Addison Wesley.

Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Smith, T.R., J. Su, A. El Abbadi, D. Agrawal, G. Alonso, and A. Saran, 1994. Computational modeling systems. *Information Systems* 19(4): 1-54.

Stewart, J.Q., 1947. *Macrogeography and Income Fronts*. Monograph Series No. 3, Philadelphia: Regional Science Research Institute.

Tobler, W.R., 1979. Smooth pycnophylactic interpolation for geographical regions. *Journal, American Statistical Association* 74: 519-530.

Turner, A.K., editor, 1992. *Three-Dimensional Modeling with Geoscientific Information Systems*. Dordrecht: Kluwer.

von Oosterom, P.J.M., 1993. *Reactive Data Structures for Geographic Information Systems*. New York: Oxford University Press.