Computers & Geosciences Vol. 18, No. 4, pp. 401–408, 1992
Printed in Great Britain.

0098-3004/92 $5.00 + 0.00

GEOGRAPHICAL DATA MODELING

MICHAEL F. GOODCHILD
National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA 93106, U.S.A.

Abstract—Data modeling is defined as the process of discretizing spatial variation, but may be confused with issues of data structure, and driven by available software rather than by a concern for accurate representation. We review the alternative data models available in spatial databases, and assess them from the perspective of accurate representation of geographical reality. Extensions are discussed, particularly for three dimensions and time dependence.

Key Words: Geographic Information System, Data model, Spatial database, Accuracy.

INTRODUCTION

Tsichritzis and Lochovsky (1977, p. 21) define a *data model* as "a set of guidelines for the representation of the logical organization of the data in a data base . . . (consisting) of named logical units of data and the relationships between them." With few if any exceptions, the world which is represented in a spatial database is not composed of logical units, and thus must be abstracted, generalized, or approximated in the process of creating a database. Data modeling thus plays a fundamental role in spatial databases, and controls the view of the world which the user ultimately receives. As the GIS industry matures, and questions of data structures, algorithms, and functionality become standardized, the critical issue of data modeling will become more and more important, both directly and indirectly through the role that it plays in such concerns as accuracy. Ultimately, a GIS will be successful only if it can present the user with an accurate view of the world; to do so requires both efficient access to a database, and the use of accurate data models. Moreover different forms of analysis and exploration of the same area will likely require different data models, depending on the form of approximation adopted in each.

The purpose of this paper is to examine the issue of spatial data modeling not from the perspective of alternative data structures, but as a process of representing geographical reality. We argue that the existence of alternative data models is one distinguishing feature of spatial databases, and creates the need for this distinct perspective. Too often the choices between data models are presented as choices between data *structures*, or specific arrangements of records and linkages within the database. If a data model consists of "named logical units of data" and such logical units are abstractions or approximations of geographical reality, it follows that one data model is not obtainable necessarily from another, because each may approximate reality in different ways.

To illustrate the distinction being drawn here between data models and data structures (compare, for example, Peuquet, 1984), consider a simple raster in which each pixel has associated with it an integer representing a census tract number. These census tract numbers point to the rows of a table containing socioeconomic data. We refer to this loosely as a raster data structure. Now suppose that a raster/vector conversion algorithm extracts the boundaries of each tract as polygons composed of vertical and horizontal pixel edges, and that each polygon now points to the same census tract table. We refer to this loosely as a vector data structure. Although the structure has changed, the information it contains is the same, as we have merely rearranged the components of the data model, and not changed the manner in which the model approximates reality. Perception of spatial variation is an important criterion in the development of data models for maps, whereas the selection of data models for spatial databases is likely to be guided by different objectives.

For the purposes of the paper we use the term *geographical reality* to refer to empirically verifiable facts about the real world. Those facts may not be certain; in practice, many of the relevant definitions include substantial uncertainty, as for example in the land use class "urban". A data model is a limited representation of reality, constrained by the finite, discrete nature of computing devices; the term *discretization* conveys much the same meaning as data model in this context. In many situations the relationship between reality and database is complicated by the interposition of a map or analog store with its own data model. Filtering then takes place both between reality and the analog store, and between the analog store and the database. The data models available to analog maps are more limited, as they are constrained by the technology of paper and pen (Goodchild, 1988b), so the double filtering which takes place makes it even more difficult to present the user with an accurate view of reality.

The paper is organized as follows. We first discuss the nature of geographical reality, and the subsequent sections review alternative data models. Extensions to three dimensions and time dependence are discussed, and the paper ends with some implications with regard to data accuracy and error modeling.

GEOGRAPHICAL REALITY

One way of defining the fundamental element of geographical information is as the tuple $T = \langle x, y, z_1, z_2, \ldots, z_n \rangle$, giving the values of a set of n *spatial variables* at the location (x, y). We allow the variables z to be of any data type: binary, nominal, ordinal, interval, or ratio. For variables measured on discrete scales, note that it is possible always to transform to a space of binary variables, or to collapse many variables to one. We assume each z to be single-valued at any location, thus excluding over-folded geological structures, which must be treated as three dimensional. We assume that the tuple is verifiable empirically, for example by visiting the location (x, y). Later in the paper we discuss extension to the tuple $\langle x, y, h, t, z_1, z_2, \ldots, z_n \rangle$ where h is the vertical dimension and t is time.

Because x and y are continuous, the number of tuples is infinite. Thus data modeling can be seen as the process of reducing the number of tuples required to represent reality to some finite set small enough to be accommodated within the constraints of a digital store. We refer to the infinite set of tuples as *a field.* x and y also are continuous in analog map stores, but here the problem of data modeling is to determine effective ways of representing the variability of z_1 through z_n. For example, it is desirable to determine ways of representing as many real variables as possible in a single mapped variable, through transformations $f(z_1, z_2, \ldots, z_n)$.

The variation of z_1 through z_n may be discrete or continuous in x and y. For parameters such as topographic elevation the data model may assume continuous variation, that is the absence of cliffs. However, geographic surfaces typically do not have derivatives (or tangents) which are everywhere well defined.

Spatial autocorrelation (Cliff and Ord, 1981) plays a key role in the task of discretizing spatial variation. We observe in general that the similarity between the variables in the tuples $T_1 = \langle x_1, y_1, z_{11}, z_{12}, \ldots, z_{1n} \rangle$ and $T_2 = \langle x_2, y_2, z_{21}, z_{22}, \ldots, z_{2n} \rangle$ increases as the locations converge. Two general strategies for discretization emerge from this observation. The *sampling* strategy exploits spatial autocorrelation by assuming that (x_1, y_1) and (x_2, y_2) must be more than a certain minimal distance apart before the associated tuples are substantially different. The *piecewise* strategy assumes that the plane can be partitioned into homogeneous, simply connected *regions,* with variation within each region described by some simple function.

Many data models are based on discrete *objects* located in the plane, allowing spatial variation to be represented by a set of tuples $\langle i, a_1, a_2, \ldots, a_m \rangle$ where i is an object and a_1 through a_m are the object's attributes. Location is described by a set of tuples $\langle x, y, o_1, \ldots, o_i, \ldots \rangle$, where o_i is a binary variable indicating the presence or absence of object i at location (x, y). The next section reviews a number of such object representations. However in most situations objects are generalizations or approximations of variation and poorly defined. For example a soil map shows the variation in soil type in an area by defining a set of nonoverlapping, space-exhausting area objects. But the locations of the boundaries are not well defined, and soil type is only approximately homogeneous within each area (Mark and Csillag, 1989; Fisher, 1989). Thus neither set of tuples may be empirically verifiable—we cannot confirm that location (x, y) is within a given object, or that all points within the object have the given attributes.

Other objects such as benchmarks and buildings may be comparatively well defined. Consider the infinite field of tuples $\langle x, y, o \rangle$ where o is a binary variable, value 1 if (x, y) is in the State of California, 0 otherwise. In this instance the object is better defined, but there are locations and levels of accuracy at which it is impossible to determine o without ambiguity, for example along the coastline. In effect, every representation of geographical reality based on discrete objects is approximate to some degree.

This object/field dichotomy is a long-standing issue in cartography (see, for example, Robinson and others, 1985, p. 108) and in the debate over cognitive understanding of geographical space (Peuquet, 1988). It emerges here in the relationship between the contents of the database and the geographical reality which they represent.

MODELS OF FIELDS

In this section we examine the alternative data models which have been exploited in spatial databases. Most of the discussion concerns the representation of a single variable, but multivariate issues are included at several points.

Piecewise models

Piecewise models partition the plane into simply connected regions, with variation described by a simple mathematical function in each region. Each location is assigned to exactly one region. Furthermore there exists at least one path between any pair of locations within the same region which is itself wholly within the region. Regions therefore may contain other regions, but may not be disconnected into islands. Many GIS data models implement the concept of a complex object, and thus allow the user to create a superregion as a union of several simple regions.

A number of forms are assumed for the function describing variation within each region.

Constant. In the simplest situation the value of the variable is constant within each region. The number of possible values of the variable now is finite, at most equal to the number of regions, and the model therefore places no restrictions on the variable's data type. In some situations the regions are defined by the variable itself, by locating boundaries in areas of particularly rapid change, allowing the model to approximate what is in reality continuous variation (Mark and Csillag, 1989). In these situations a second step of discretization is necessary in order to represent the continuous curves of the boundaries in digital form, usually by selecting a finite set of points and connecting them by straightlines. This form of discretization is convenient merely as the object being represented has no existence in reality.

In other situations the boundaries will have been defined by some process which is independent of the variable itself. For example, much socioeconomic data are discretized by using *reporting zones* with boundaries which follow streets, rivers, railroads, etc. In this situation also the discretization of boundaries almost always is by points connected by straightlines, although the nature of the phenomenon being represented may suggest better alternatives. A number of systems allow arcs of circles as well as straightline segments.

Constant piecewise approximations generally are used to describe spatial variations in soils, land use, land cover, and many other biophysical variables. The identical data model is used for much socioeconomic data, although homogeneity within zone may not be assumed in analysis of such data. In the biophysical situation it is usual for each variable to produce a unique discretization, but in the socioeconomic situation the set of boundaries usually is common to many variables.

Consider the set of biophysical variables $z_1, \ldots, z_n$, each with its own associated set of regions $\mathbf{R}_i$. After discretization of variable i, each region is assigned an attribute $a_i \in S_i$ measured on a discretized version of the scale of measurement of z_i, where S_i denotes the discrete domain of the ith attribute. At any point (x, y) we can identify the attributes $\langle a_1, \ldots, a_n \rangle$ assigned to the regions which the point occupies in each discretization. The concatenated attributes at (x, y) are an element of the Cartesian product $S_1 \times S_2 \times \cdots \times S_n$, and the associated regions $\mathbf{P}$ are the familiar product of topological overlay of the sets of regions $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_n$.

The regions $\mathbf{P}$ can be obtained in two clearly distinct ways. First, each variable i can be discretized, and the resulting regions overlaid. Alternatively we might attempt to discretize the multivariate space defined by $z_1, z_2, \ldots, z_n$ directly. In the space defined by the variables $z_1, \ldots, z_n$ the first situation would result in a partitioning by hyperplanes perpendicular to the axes; in the second, there would be no constraints on the geometry of partitioning. The terms ITU (integrated terrain unit) and LCGU (largest common geographical unit) are associated primarily with the first approach. As such, the debate on whether to use ITUs or independent discretizations as the basis for multivariate spatial databases (Burrough, 1986, p. 4) is essentially an issue of data structure rather than data model, and its resolution depends on the comparative costs of data processing and storage rather than on the accurate representation of reality.

Linear. Now suppose that the variation within each region is described by a plane, or linear function $a_0 + a_1x + a_2y$. We require the scale of measurement of z to be continuous. The TIN model (Burrough, 1986) is a particularly simple form of linear piecewise approximation where all regions are triangles, and nodes are restricted to triangle vertices. The major reason for adopting triangles is that it is easy to ensure continuity of elevation across triangle edges.

Although there are many examples of spatial variables measured on continuous scales, the TIN model is applied only in the situation of topographical elevation. Many landscape-forming processes are responsive to gradient, and tend to produce terrain with substantial areas in which gradient is constant, although glacial processes are a notable exception (Mark, 1979). The TIN model allows the sizes and positions of triangles to adapt to the complexity of the terrain, with smaller triangles in rugged areas. Recent versions of TIN models allow the user to define TIN vertices interactively at critical points on the surface, and to position triangle edges along lines of observed discontinuity of gradient (McCullagh, 1988).

Higher order functions. More complex functions offer the possibility of more accurate representation of variation within each region. Akima (1978) has described the use of quintic polynomials within each triangle of a TIN, with the advantage that gradients can be made continuous across triangle edges. However, although this may by useful in producing visually acceptable contours from irregularly spaced point data, it is less so as an accurate depiction of terrain. Discontinuities of slope are widespread on real topography, and can be modeled with planar TINs by aligning triangle edges with observed ridges and valleys.

Contours

If the scale of measurement of z is continuous, and if the variable is autocorrelated strongly spatially, then the set of points $\langle x, y \mid z \in c_1, c_2, \ldots, c_m \rangle$ defines a set of *contour lines*. Conventionally we assume that the contoured values $c_1, \ldots, c_m$ are evenly spaced along the domain of z. The set of points forming each contour line can be discretized further as a finite set connected by straightlines. The resulting contour lines partition the plane into regions, in each of which the value lies between two consecutive contoured values. However, the regions produced in this way have certain distinct characteristics imposed by the continuous nature of z. The associated boundary network has no nodes of valency greater than 2, and two regions can be adjacent only if the associated contour intervals share a common value.

The popularity of the contour model as a way of depicting topography on analog maps suggests that it is optimal for this particular variable and the technology of cartography. It undoubtedly is an efficient way of communicating information on the spatial variation of a continuous variable to the user, and is reasonably successful for visualizing two variables. However, as a method of discretization in the relatively unconstrained environment of spatial databases, it has a distinct disadvantage, as the level of approximation differs dramatically across the plane, being maximum on contour lines and minimum midway between them. The accuracy at any point has no relationship to the phenomenon, being controlled entirely by the arbitrary selection of contoured values. Nevertheless digitized contours continue to be a readily available source of topographic information, which perhaps is the most convincing example of the filtering effects of analog map data models on spatial databases.

Sampling

Spatial variation can be discretized by capturing the value of a variable at a finite set of points. A *raster* results if the points are uniformly spaced, whereas irregular sampling may be more efficient if the density of sampling can adapt to the local degree of variability. The TIN model provides a simple way of interpolating between irregularly spaced sample points in the situation of a continuous scale of measurement. In general, however, sampling imposes no constraints on scales of measurement or on the range of values possible at each point.

PLANAR ENFORCEMENT

Thus far we have been concerned with the modeling of one or more variables whose values are defined everywhere in the plane. In many situations it is more convenient to view reality as an empty plane littered by objects, which may be points, lines, or areas. Any location (x, y) is either empty or occupied by one or more objects. Each object has a set of associated attributes which serve to differentiate it from other objects. We already have commented that the quality of definition of such geographical objects is highly variable, and almost always less than perfect.

We use the term *planar enforcement* to refer to the rules used in converting this form of representation to a single-valued function defined everywhere. Planar enforcement occurs at many points in spatial data handling, and we consider three in particular, with associated examples.

Consider the task of digitizing region boundaries from a map. This operation creates a set of line objects littering the plane, or "spaghetti". In order to create a set of regions in which every location has a single value it is necessary to first "snap" lines at junctions and remove overshoots, and then to obtain the attributes of the regions thus formed in some consistent way. In some situations this latter step is achieved by assigning the attributes of an arbitrarily located point to the containing region, and in other situations by assigning attributes attached to each side of each line object. The GIS industry may refer to the whole operation as "building topology", and the term "cartographic" may be associated with the view of the world as a plane littered by (unrelated) objects.

Spatial interpolation is the term usually given to the task of computing a complete, continuous surface from a set of sample points (Lam, 1983). In this instance a value is obtained everywhere in the plane from attributes attached to a finite set of point objects. Spatial interpolation can be defined for any scale of measurement, but is applied usually when the scale is continuous.

Now consider a set of objects lying in the plane with the attribute "woodlot". We can assume safely that these objects will not overlap, but if they do there is no particular problem in assigning the same attribute to their intersection. But suppose the objects represent forest fires and the attribute is date. Because a location may have burned more than once it is not obvious immediately how the tuple $\langle x, y, z \rangle$ might be defined at each point. One alternative would be to make z a count of fires; another would make z the date of most recent fire, with a special code for unburned regions.

VARIATION ON NETWORKS

Thus far we have been concerned with the representation of variation over the plane, whether viewed as an infinite field of tuples or as a space littered with discrete objects. However neither of these views is consistent particularly with GIS applications in transportation or surface hydrology. In these areas data modeling requires two mostly independent stages. The first models the network as a collection of objects embedded in the 2-D plane. The objects are typically nodes and links. In many applications it is important to distinguish between links which cross geometrically and links which intersect at a node, to allow for grade separations. In this sense planar enforcement may be inappropriate for these networks.

The second stage models the variation of phenomena along the network. Attributes may be associated with points, such as bridges or houses, or variation may be modeled by piecewise discretization. For example, the variation of pavement quality over a highway network might be modeled by defining segments with homogeneous quality. Variation in elevation of railroad networks may be modeled by defining segments of constant gradient. We use the term *segment* to refer to a discrete element of a network.

In the first stage of discretization, locations are defined in the (x, y) plane. In the second stage, however, locations are defined more conveniently by a pair of the form ⟨link, offset⟩ such as street address, or the ⟨route, milepost⟩ addressing system used by railroads. In summary, the first stage requires the definition of a set of line objects; a location (x, y) may or may not be occupied by one or more objects. In the second, an infinite set of tuples $\langle l, o, z_1, z_2, \ldots, z_n \rangle$ is defined over the network, where l defines a link and o an offset distance from the origin of the link.

Just as the plane allows independent discretizations for each variable, it should be possible to define independent segmentations of a network without repeating the first stage of discretization in every situation. Unfortunately many current GIS products do not allow this; instead, both levels of discretization must be collapsed to one. Because link objects are allowed only homogeneous attributes in these systems, it is necessary to create nodes wherever a change of attributes occurs, in effect forcing the equivalent of an ITU strategy. For example a node must be positioned at every point event or change of attributes on a rail network, including stations, switches, tunnels, bridges, etc., leading to almost endless proliferation of link objects.

In essence, transportation networks are not sets of linear objects littering the plane, but 1-D addressing systems embedded in 2-D space. The values of spatial variables are defined only on the network, and not in the intervening spaces. Similar issues of multiple levels of discretization exist in 3-D spaces and in the time-dependent case (see next). It also is possible to determine examples of spatial variables whose values are defined only at points.

CLASSES OF OBJECTS

We have seen how spatial databases reflect two different views of reality—as infinite sets of tuples approximated by regions and segments (the field view), and as planes littered with independent objects (the object view). The concept of an object arises in both situations, but in the first the area and line objects representing regions and segments cannot exist independently, but instead must partition the plane and the network respectively, and must be organized into well-defined, planar-enforced layers. The rules affecting the behavior of objects in the two views therefore are different.

The term "object-oriented" (OO) has received attention recently in the GIS literature (Egenhofer and Frank, 1988a, 1988b) as many of the computer science concepts of object-oriented programming and databases have stimulated discussion in a spatial context. The OO notion of object identity clearly is more compatible with the object view of reality than with the field view, and the systems currently being marketed as "object-oriented" rather than "layered" seem to be aimed at those applications in which the object view is more acceptable. Kjerne and Dueker (1988) have discussed the implications of the OO concept of inheritance in survey data, whereas Armstrong and Densham (1989) discuss the implications of the OO concept of encapsulation for spatial analysis and modeling. It will be some time before the full impact of OO on GIS becomes clear.

In summary, we define five classes of objects in two groups:

Field view: region, segment
Object view: point, line, area.

The regions and segments in the field view must be grouped to partition collectively the plane or network, but the classification of objects in the object view is more flexible.

The relationships between objects fall into three types:

(1) Those which are necessary for the definition of objects, for example the relationships between points which define a line;
(2) Those which are computable from the geometry of objects, for example the "contains" relationship between point and area, or the "crosses" relationship between two lines;
(3) Those which are not computable, for example the "intersects" relationship between two roads.

Finally, we introduce the concept of an *object pair* (Goodchild, 1988a), a virtual object created from the relationship between two simple objects, which may itself have attributes. More formally, an object pair is the tuple $\langle i, j, a_1, a_2, \ldots, a_m \rangle$ where i and j are two objects of the same or different classes and a_1 through a_m are attributes. The attributes of an object pair are not associated normally with any simple object. For example, the relationship between a point object "sink" and another point object "spring" can have attributes of distance, flow, flow-through time, etc., but has no physical existence as a defined spatial object. Object pairs are important for modeling various forms of spatial information, and are implemented in ESRI's ARC/INFO, for example, as the "turntable" in the NETWORK module.

ACCURACY

General strategies

If the purpose of a spatial data model is to represent an infinite number of real tuples using a finite number, then accuracy is clearly an important issue in selecting between alternatives. Unfortunately the appropriate objective function to use in defining accuracy depends on the use to which the database will be put. It is tempting, for example, to assume that the appropriate measure for topographic elevation is the mean absolute difference between real and modeled height at a randomly selected point, but this measure is less useful than accuracy of aspect to someone concerned with modeling surface flow directions, which depend only on aspect.

In many situations there are several stages of discretization between reality and the database. For example we might model elevation by measuring spot heights photogrammetrically, drawing smooth contours, digitizing them, and finally building a TIN model from the digitized contours. Information is lost at each step, and in some steps spurious information is introduced, particularly in the interpolation of smooth contours and the construction of the TIN.

Two general strategies seem appropriate. First, it is desirable to minimize the loss of information between reality and the database, by minimizing the number of discretizations and manipulations. In the example the data are discretized probably already sufficiently as spot heights, and the use of contours could be avoided altogether. Second, it is desirable as far as possible to serve diverse needs by creating several distinct views of the database. For example, the contour map could be derived as a view of the (spot height) database, rather than as a step in the complex input process. The ability to create customized views for different purposes should be one of the major advantages of a database approach. In principle, such cartographic devices as scale and generalization, which reflect different views rather than properties of reality, should be treated as far as possible as attributes of a given view of the database, rather than as attributes of the database itself.

Objects vs fields

Although field and object views of reality are to some extent alternative perspectives, we have argued in this paper that the tuples of a field are empirically verifiable, whereas it is possible only to confirm the existence of an object imperfectly. For this and related reasons Goodchild (1989) has argued that error is treated more appropriately from the perspective of field rather than object. The objects compiled by cartographic processes may be stripped of any information on which a useful model of error might be based, by processes such as low-pass filtering of boundaries, and deletion of small regions.

The distinction between fields and objects in modeling error is best illustrated with the example of contours. We regard a given set of contours as a single sample from a population of possible sets, and require that there be no difference between two sets other than the effects of error. It would be unacceptable, for example, if one set were substantially longer or more wiggly than the other. Then it seems to be impossible to write down a stochastic process which would take one such sample and produce another by any form of distortion. For example, adding a Gaussian error in x and y to a randomly selected set of points along each contour clearly would lengthen the contours, so the two versions would differ significantly. On the other hand, it is comparatively simple to take the field of elevations from which the contour objects were created, add a suitably spatially autocorrelated error field, and obtain a second set of contour objects.

THE RASTER/VECTOR DEBATE

The discussion to this point has avoided deliberately the terms "raster" and "vector", despite the fact that these may be used to characterize two distinct classes of GIS software. The first section of the paper argued that the same information could be restructured readily from a form which would be labeled broadly as "raster" to one which would be accepted as "vector". This section reexamines the raster/vector debate within the framework of data modeling established in the previous sections.

"Raster" refers to a data model based on a regular (usually rectangular) tesselation of the plane, in which all location information can be imputed from a record's sequential position, and therefore is missing from the data structure. However, the geometry of a regular tesselation can be used to model spatial variation in numerous ways. First, the value attached to a pixel is assumed to apply homogeneously to the pixel's entire area. In this situation the raster model is a special example of the piecewise constant models described previously. In the situation of a DEM the value usually is an estimate of mean elevation within the pixel, and this also is the situation with remotely sensed images. In other instances the pixel's value is the modal or most abundant class, or the class at the central point. Further ambiguity arises in the situation of a DEM, because local estimates of slope and aspect may be made by fitting a plane to a small neighborhood, implying a piecewise linear rather than piecewise constant model.

The term "vector" is similarly ambiguous from a data modeling perspective. It can indicate an irregular polygonal tesselation, with piecewise constant variation, or a TIN, or a set of contours, or an unstructured CAD file containing points, lines, and areas. Thus neither "vector" nor "raster" provide unambiguous information on how the data model reality. A mapping exists between these terms and the alternatives discussed in this paper, but it depends on conventional usage and is loose. Thus the assumption made in this paper is that "raster" and "vector" do not provide an adequate and sound basis for a discussion of data modeling.

EXTENSIONS TO THE MODEL

Complex objects

Several contemporary GIS designs include the possibility of defining complex objects as collections of simple objects. A complex object may have its own graphic transform—for example, a collection of points may be represented as a point—and its own attributes, some of which may be aggregations or means of simple object attributes. The mapping between simple and complex objects may be $n:1$ or $n:m$. The concept of a complex object is clearly incompatible with the field view of reality.

Shared primitives

In this extension an object such as a point or line may be shared between a number of objects. For example the common boundary between two polygons may follow a road: the road and the two segments of common boundary would be defined as a primitive shared between the three objects. Any update of the shared primitive would thus modify all three objects. Again the role of this extension is different in the two views of reality. In the field view all arcs are by definition shared, and it would be inefficient not to treat them as common primitives. On the other hand, sharing would not be appropriate always in the object view, even though the relevant objects might be coincident.

Inheritance and lineage

The model contains no explicit way for attributing accuracy or lineage to objects, or of propagating these attributes through GIS processes. The technology to do so at this point is limited.

THREE DIMENSIONS AND TIME

The concepts of field and object views apply equally to 3-D data (Ganter, 1989), where one can visualize space as occupied by a collection of objects, or by a single-valued function. The B-rep and SOE options, which correspond roughly to vector and raster respectively, are alternative ways of structuring information, rather than alternative models. Any point in the space may be associated with the values of one or more functions in the tuple $\langle x, y, z, z_1, z_2, \ldots, z_n \rangle$, or by the presence or absence of one or more point, line, or area objects. Alternatively variation may be defined over a surface or a set of lines embedded in the space.

Although the modeling of 3-D data presents no significant problems, and is widespread in such fields as CAD and medical imaging, it differs from the 2-D situation in the lack of analogs. There is no equivalent of the map to act as an input medium or filter. Consequently, although 3-D data models exist in the digital domain, the creation of data for them presents much more of a problem, and there is a general lack of suitable data for 3-D GIS. High priority should be given to the development of a 3-D data compilation workstation, which would allow a user such as a geologist to input evidence of various forms (well logs, seismic data, expert knowledge) and build 3-D data models through a variety of forms of spatial interpolation. The same issue is missing in two dimensions because it is so easy for data compilation to take place using the map analog.

In the instance of time, the asymmetry between time and the spatial dimensions introduces further complexity. There appear to be five major modeling options:

(1) A finite number of discrete time slices, each viewed as a field;
(2) Discrete time slices, with objects identified in each slice but with no information relating objects between slices;
(3) Discrete time slices, with static objects which are present or absent in each slice;
(4) Discrete time slices, with objects identified in each slice and linkages between corresponding objects at different times;
(5) Continuous time, creating a 3-D space in which the movement of 2-D objects is represented by 3-D objects (points become lines, lines become surfaces, areas become volumes).

DISCUSSION

In this paper we have tried to distinguish clearly between data structures and data models, defining the latter as alternative discretizations of the infinite complexity of spatial information. Data modeling in spatial databases seems to adopt two alternative views, depending on whether it regards reality as occupied by a set of single-valued functions defined everywhere, or by a set of objects. Although the data structures used in the two situations may be identical, they nevertheless represent different perspectives on reality.

The selection of data model is critical in spatial data handling, because ultimately it affects the views which the database presents to the user, and which the user judges against empirical truth. Unfortunately, selections may be limited by the set of models associated with analog maps, or by the set offered by the vendor. An effective selection between alternatives also requires a degree of understanding of the nature of the geographical phenomenon being represented, and the processes which created it. Finally, selection also affects the extent to which it is possible to model and understand uncertainty, and its propagation through the steps of analysis. Many of the supposed benefits of GIS technology—the ability to change scale and overlay, and the separation of the roles of data collection and analysis—are in some ways its greatest weaknesses.

REFERENCES

Akima, H., 1978, A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points: ACM Transactions on Mathematical Software, v. 4, no. 2, p. 148–159.

Armstrong, M. P., and Densham, P. J., 1989, Object oriented locational analysis: Proc. GIS/LIS '89. ASPRS/ACSM, Bethesda, Maryland, p. 717–726.

Burrough, P. A., 1986, Principles of geographical information systems for land resources assessment: Clarendon, Oxford, 193 p.

Cliff, A. D., and Ord, J. K., 1981, Spatial processes: models and applications: Pion, London, 266 p.

Egenhofer, M. J., and Frank, A. U., 1988a, Object-oriented databases: database requirements for GIS: Proc. Intern. Geographic Information Systems Conference: The Research Agenda. NASA, Washington, v. II, p. 189–211.

Egenhofer, M. J., and Frank, A. U., 1988b, Designing object-oriented query languages for GIS: human interface aspects: Proc. Third Intern. Symposium on Spatial Data Handling, Columbus, Ohio, p. 79–96.

Fisher, P. F., 1989, Knowledge-based approaches to determining and correcting areas of unreliability in geographic databases, *in* Goodchild, M. F., Gopal, S., eds., Accuracy of spatial databases: Taylor & Francis, London, p. 45–54.

Ganter, J. H., 1989, A comparison of representations for complex earth volumes: Proc. AutoCarto 9. ASPRS/ACSM, Falls Church, Virginia, v. 5, p. 278–287.

Goodchild, M. F., 1988a, Towards an enumeration and classification of GIS functions: Proc. Intern. Geographic Information Systems Conference: The Research Agenda. NASA, Washington, v. II, p. 67–77.

Goodchild, M. F., 1988b, Stepping over the line: technological constraints and the new cartography: The American Cartographer, v. 15, no. 3, p. 311–320.

Goodchild, M. F., 1989, Modeling error in objects and fields, *in* Goodchild, M. F. and Gopal, S., eds., Accuracy of spatial databases: Taylor & Francis, London, p. 107–114.

Kjerne, D., and Dueker, K. J., 1988, Modeling cadastral spatial relationships using Smalltalk-80: Proc. GIS/LIS '88. ASPRS/ACSM, Falls Church, Virginia, p. 373–385.

Lam, N. S.-N., 1983, Spatial interpolation methods: a review: The American Cartographer, v. 10, no. 2, p. 129–149.

Mark, D. M., 1979, Phenomenon-based data structuring and digital terrain modeling: GeoProcessing, v. 1, no. 1, p. 27–36.

Mark, D. M., and Csillag, F., 1989, The nature of boundaries on 'area-class' maps: Cartographica, v. 26, no. 1, p. 65–78.

McCullagh, M. J., 1988, Terrain and surface modelling systems: theory and practice: Photogrammetric Record, v. 12, no. 72, p. 747–779.

Peuquet, D. J., 1984, A conceptual framework and comparison of spatial data models: Cartographica, v. 21, no. 4, p. 66–113.

Peuquet, D. J., 1988, Representations of geographic space: toward a conceptual synthesis: Annals, Assoc. American Geographers, v. 78, no. 3, p. 375–394.

Robinson, A. H., Sale, R. D., Morrison, J. L., and Muehrcke, P. C., 1984, Elements of cartography (5th ed.): John Wiley & Sons, New York, 544 p.

Tsichritzis, T. C., and Lochovsky, F. H., 1977, Data base management systems: Academic Press, New York, 388 p.