TEMPORALITY IN SPATIAL DATABASES

Marc P. Armstrong
Departments of Geography and Computer Science
The University of Iowa
316 Jessup Hall
Iowa City, IA 52242
BLAMGPD@UIAMVS.BITNET

ABSTRACT

Time and the examination of change play important roles in spatial decision-making. The incorporation of time in spatial databases, however, has been largely performed on an ad hoc basis. In this paper I review approaches to handling temporal data, and develop general models for incorporating temporality in spatial databases.

INTRODUCTION

Most social constructions of time mark its passage along a linear dimension (Lauer, 1981; Horwich, 1987). This enables time to convey a sense of the past, present, or future, and allows events to be placed along an asymmetrical forward-moving temporal continuum. Causation and explanation depend upon this property of temporal asymmetry, and as a result, temporal aspects of database management are important to the successful implementation of computer systems designed to support decision-making (Clifford and Ariav, 1986). Yet this important aspect of spatial database design is often overlooked. The aim of this paper is to provide a framework for incorporating time, and its consort, change, in spatial databases. I first provide a brief overview of spatial database organization strategies, and then discuss temporality, and how temporal data have been treated in spatial databases. In the final section of the paper a framework for handling temporal data is described.

SPATIAL DATABASES

Spatial database structures provide a means for describing the metrical, topological, and attribute characteristics of objects. In grid cell databases a rigid geometrical structure is imposed upon the environment. The metrical and topological configuration of the grid remains unaltered over time, and as the environment changes, the grid filter captures the attributes in each cell. Many commercial systems, however, use vector models to describe objects because they enable fidelity to be maintained between objects and their digital representations, and allow the metrical properties of entities to be varied without disrupting topology. This property of vector models is useful for handling temporal change, because the form of an object can be represented at different time intervals.
Attribute classification schemes also affect change assessment (Anderson, 1980: 103). In Figure 1, when low categorical precision is used on the left, change is not observed between times 1 and 2. If higher precision is used, however, two types of change are observed: 1) a change in category without a change in morphology; and 2) a morphological change (decomposition of the original polygon) resulting from changes in category. Temporality is important in spatial databases because metrical, topological, and attribute properties of entities can change either together, or independently.

![Low Categorical Precision](image1)

![High Categorical Precision](image2)

Figure 1. Effects of attribute classification on change.

**TIME AND GEOGRAPHICAL INQUIRY**

Time has played an important role in the development of geographical knowledge in areas such as diffusion, migration, and time geography, while researchers such as Getis and Boots (1978), and Bennett (1979) also have incorporated temporal components into spatial models. Other researchers have explored spatial processes through dynamic data display (see Calkins, 1984) and change detection techniques (e.g. Jensen and Toll, 1982).

Given the importance of time in geographic research, it was inevitable that a temporal component would be added to data organization strategies. More than twenty years ago, Berry (1964) constructed a geographic matrix with places and their characteristics forming columns and rows. But because variation occurs across both space and time, Berry extended the matrix in a third dimension to incorporate temporality. In its changed form, multiple discrete time "slices" were introduced. This view of data organization became a dominant data organization paradigm in geography. Haggett, Cliff and Frey (1977:15), for example, show a geographical data cube in which spatial objects and attributes comprise two dimensions, and time, the third.

Time also has been examined by GIS researchers. Dangermond (1983), for example, employs a figure similar to Haggett, et al. (1977) to illustrate a temporal
component in GIS databases. Basoglu and Morrison (1977) recognized a hierarchy in data organized for studying the evolution of county boundaries from 1790 to 1970. In implementation, each line segment record is owned by a state and county, and is linked to a date file which describes a valid time interval for the line segment.

More recently, Burrough (1986) asserts that the dynamic nature of environmental features is poorly handled by conventional GIS techniques. Chrisman (1983) argues that what may seem to be logical errors in GIS analyses often can be attributed to mismatched information from various time periods. Space, time and attribute information, therefore, are seen by Chrisman to interact, and their quality must be accounted for if a GIS is to remain useful. In other work relating to data quality, Kennedy-Smith (1986) suggests that a "delta file" be maintained. Such a file consists of differences (deltas) between various editions of a file, and helps to maintain compatibility among databases. In a delta file entities may be deleted, may undergo a position change because of either a resurvey or adjustment, or may undergo an attribute change.

Clearly, the incorporation of processes into the realm of geographical inquiry is well established. Yet the incorporation of these ideas into functional systems has been performed inconsistently, and the level of conceptual development is inadequate for modern GIS applications.

**Temporality in Database Management**

Early studies of temporality in the database management literature (e.g. Bubenko, 1977) draw a distinction between physical time, which refers to transactions (e.g. update) and their registration, and logical time, which refers to the demarcation of events as they occur along a temporal continuum. In a synthesis and extension of this earlier work, Snodgrass and Ahn (1987) describe four ways in which temporal data have been organized, and introduce several new concepts of time for databases.

**Static Database.** In static databases only the current state of the database is recorded; when new data are added, all changed data are discarded. Because of the resulting inability to reconstruct temporal sequences, this approach is ill-suited for most GIS applications.

**Static Rollback.** In this approach, each past state is stored in time-indexed form. It is then possible to examine time-slices (rollbacks), but changes can be made only to the most recent static entry in the database. Another problem with the static rollback approach is that transaction time is used, rather than the time at which each event occurs. Furthermore, the approach is inefficient if only small portions of the database are changed in an update operation, because all unmodified data are stored in duplicate for each time slice.
Historical Database. In historical databases, a historical state of each entity is kept. In this view, the concept of valid time becomes important. Valid time, specified either in point, or interval format, is related to the occurrence of events and designates a period of validity for an entity, relationship or attribute in a database. Also, in this approach it is possible to update any, rather than only the most recent, version of the database.

Temporal Database. In temporal databases, both transaction and valid times are maintained. Snodgrass and Ahn (1987) also discuss one additional representation of time: user-defined. This is useful for denoting events that may have lagged effects. For example, in a GIS database it may be important to encode information obtained from an aerial photograph for a data layer:

1) The time at which the imagery was sensed is the valid time of the map.
2) The time when data are digitized for entry into a GIS is the effective (user-defined) time of the layer.
3) The time of supplementing the database in an update operation is the transaction time of the layer.

In the above discussion, the issue of temporal resolution has been ignored. Point and interval representations of events or states are fundamental units used to specify time in a database. De et al. (1987) enumerate these temporal primitives. Another resolution issue involves the choice of an appropriate class for representing temporality. Furtado and Neuhold (1986) use a dual approach; time is separated into: Time_of_Day (hour, minute, and second), and Date (year, month, and day). The latter formulation is likely to be used in most geographical database applications.

TEMPORAL DATA IN SPATIAL DATABASES

The framework developed in this section is derived from the preceding discussion about temporal data. The addition of a spatial component to temporal data management concepts, however, adds complexity because entities may change location and shape.

Spatial Considerations

In this section, several components of change are identified, and although each may operate alone, more commonly, they operate in concert.

Geometrical transformation. At a primitive level, an object can translate or rotate through space without altering its form. Scaling results in a change in size, but not shape of an object, while other transforms, or combinations, may create changes to size, shape and location.
Accretion and erosion. In this case, processes with stochastic components operate on an object to cause change (e.g. conversion from field to forest).

Scale and dimensional change. Entities represented as points at time T1, may grow and at a given scale at time T2, may be represented by a line or area. Furthermore, the dimension of an object may require alteration at different scales to permit cartographic representation.

To compound complexity, as an entity’s shape and location change, its topological relationships also vary. Finally, note that attributes of entities remain invariant in this discussion. As attributes vary, however, change must be denoted irrespective of spatial configuration.

Temporal Considerations

A number of temporal issues also must be considered:

Explicit or implicit denotation of events. Temporality may be specified using regular or irregular intervals. When irregular sampling is used, it is implied that a significant change has occurred in morphology, topology, or attribute. On the other hand, if a regular sampling interval is used, it is assumed that the interval will not mask or alias change. In either case, however, we must make assumptions about the form of change that occurs between samples.

Duration or instancing. This issue centers on whether it is important to maintain a record of the duration of the status of an entity, or denote the occurrence of events. The latter case may serve well for ephemeral phenomena, or for entities for which raw counts are required (e.g. traffic counts).

Implicit or explicit duration. If phenomena are continuous, and a change in status is recorded, duration can be computed on the basis of the temporal difference between the previous and current state. If discrete phenomena are considered, both the starting and ending of the event must be recorded. This distinction is partly conditioned by the theme and resolution of data.

Priority of space, time, or entity. In a data organization strategy, priority may be assigned to a method of access. A particular application may require that temporal or spatial primacy be selected.

These issues are embodied in the following approaches to organizing temporality for areal data. In the case of grid cells, or nearly invariant polygonal layers such as census tracts or counties, the process of accounting for polygonal evolution is straightforward. In other instances, however, the accounting web becomes more complex because polygons may undergo a complete geometrical, topological and attribute transformation, which may erase links to any prior polygonal
tessellation. One way to deal with such complexity is to store layers in terms of what has changed from a previous time in the spirit of the delta file of Kennedy-Smith. But this may be inappropriate for many applications.

**SPACE-TIME DATABASE ORGANIZATION FOR AREAL DATA**

The spatial database views described in this section conform to the historical and temporal descriptions of Snodgrass and Ahn (1987). Although static or static rollback spatial databases can be constructed, they will not be considered because of their limited utility. The discussion focuses on developing general models for representing space-time databases which are presented as generalized entity-category-relationship (ECR) diagrams. The ECR approach to database design (Navathe, et al. 1986) is a direct extension of the entity-relationship modeling strategy devised by Chen (1976). The ECR approach, however, introduces categories and permits the specification of structural constraints to improve semantic modeling capabilities.

**Grid Cell Tessellations**

In grid cell representations, assuming that grid size remains constant, temporality can be handled by accumulating data layers (Figure 2). This approach is costly, however, because even cells that do not change are stored at each time period. An alternative approach encodes row and column locations, followed by the attributes associated with each cell. Each cell now has an independent capacity to take values for different times because we are not forced to deal with entire layers (Figure 3). This process can be made efficient by sorting entries if it is assumed that the most current values will be used most often, and that frequency of access to layers will decrease over time. Change detection, of course, is still dependent upon the sampling interval, and attribute resolution used.

**Polygonal Tessellations**

When polygon (or vector) structures are employed, potential complexity is increased. The main source of complexity stems from the absence of spatial uniformity. The combinations of metrical (morphology), topological, and attribute change that can occur are:

1) No change;
2) Morphology change, topology and attributes retained;
3) Morphology and topology change, attributes retained;
4) Attributes change, morphology and topology retained;
5) Attributes and morphology change, topology retained;
6) Topology change, morphology and attributes retained;
7) Topology and attributes change, morphology retained;
8) Complete morphological, topological and attribute transformation.
Figure 2. Temporally stratified layers.

Figure 3. ECR diagram for a temporal grid.

Figure 4 illustrates these configurations. Note that if the background is treated as a separate polygon, then morphological change would occur in some of the examples for which no morphological change is noted.

**Snapshot View.** This view is applied to polygonal layers with nearly invariant borders (e.g. U.S. counties since 1960). When this view is used, areal objects such as polygons or cells have valid time-stamped attributes (e.g. Census of Population on 1 April 1980). The time-stamp can occur at regular or irregular intervals, and can lead to the insertion, deletion, or modification of information. The utility of this representation can be extended if we relax a strict hierarchical assumption about database organization. In this case, each location has many dates, and each date has many locations (Figure 5). It now becomes possible to retrieve temporal information for, say, a given polygon, and also to retrieve all polygons associated with a given time. When the snapshot view is applied, duration can be denoted implicitly or explicitly. In the first case, continuous phenomena are assumed. Duration is calculable.
in this context, given by: \( D = T(n) - T(n-1) \). For a fixed temporal sampling strategy, duration can be estimated by simple enumeration of a counter. For a known (fixed) number of changes, an entity with no mobility can be represented with a single record:

| Birth | Event1 | Event2 | Event3 | Death |

Figure 4. Depiction of change categories.

Figure 5. ECR diagram for snapshot view.
Without a fixed number of occurrences, this structure is inappropriate, because of problems with repeating groups.

When change is denoted explicitly, the duration is included, or made calculable on a given record, thereby reducing I-O requirements (Figure 6).

![ECR Diagram](image)

**Figure 6. Explicit denotation of duration.**

In this instance, for an object the end of a state may not correspond with the start of another state; an intermediate status can exist that is not recorded.

**SUMMARY**

The incorporation of temporality in spatial databases engenders considerable complexity in data representations. I have attempted to delineate some of the major issues associated with time and change for areal data, drawing on the database management and spatial analysis literatures for inspiration. ECR diagrams depict selected issues. Note however, that points and lines have unique temporal characteristics which are beyond the scope of this discussion.

**REFERENCES**


