Frank, A. U., and T. B. Buyong. "Geomertry for Three Dimensional GIS in Geoscientific Applications." In *Nato Atw on 3d Modelling with Geoscientific Information Systems*, edited by K. Turner and et al., 233-57: Kluwer Academic Publishers, 1992.

Andrew U. Frank
Taher B. Buyong
National Center for Geographic Information and Analysis
and
Department of Surveying Engineering
University of Maine
Orono, ME 04469
bitnet: FRANK@MECAN1
bitnet: TBUYONG@MECAN1

Abstract: The geosciences are confronted with data handling problems which involve highly complex three dimensional objects. This complexity demands advanced analytical systems similar to the conventional two dimensional geographic information systems. Building such a system requires a thorough understanding of the geometric concepts used in the geosciences, and the applicable data models and data structures. Geometric concepts deal with geometry as perceived in geosciences including the applicable spatial and temporal reasoning. Geometric data models are the formalization of these concepts suitable for computer implementations. For each data model, there are several data structures that can be used to implement them. The concern for data structures is to provide the operations as required in the data models in the most efficient way. This is usually done by varying the implementation strategies to suit the applications.

1. Introduction

Discussions of data structures to model geometry for geographic information systems (GIS) have progressed considerably over the last 15 years. The problem is essentially to model geometric concepts describing reality using a computer system. This does not seem to be difficult. However, research and development efforts of recent years have often contributed more to our understanding of the depth of the problem than to a final solution.

Initially, the problem was considered one of optimal data structures on a very low level, close to the organization and operations of basic computer hardware. Examples of this level of discussion can be found in [Dutton 1978], which is representative of the research at that time. Most of the work treated the two dimensional, plane geometry, but there was some consideration for applications that would require three dimensions [Thomas 1978; Tipper 1979]. Research during this time was concerned with the computer aided treatment of cartographic data and industry produced computer assisted map maintenance systems. At the same time, there were papers discussing the analytical capabilities a GIS could offer to geography and other geosciences [Driel 1975; Weller 1975]. This appeared to be extremely attractive, but revealed some problems with the treatment of geoscientific data as represented by traditional maps.

Data structures to represent geometric data were also needed in CAD/CAM (computer aided design/computer aided manufacturing) systems. They started as systems to facilitate the production of paper drawings (CAD) but with the promise of extending further into the design and manufacturing process. Similarly as in GIS, the limitations of representing geometric concepts with the tools of traditional drawings became apparent. [Requicha

1980] is a frequently referenced paper that gave a survey of suitable data structures to model volumes. Two major structures appeared, i.e. constructive solid geometry and boundary representation, and for each, a number of variants for implementation were elaborated in the following years [Ansaldi, De-Floriani, and Falcidieno 1985; Gargantini 1982; Meagher 1982; Weiler 1985].

Understanding the limitations of computer assisted map maintenance systems pointed the way to data structures which represent geometry, not the map image of geometric phenomena. Frank [1984] argued for a clean differentiation between systems that deal with data directly representing some geometric reality and systems that deal with map representations. Only the first can support sophisticated geometric analytical functions, whereas the latter facilitate the production of traditional maps that can be analyzed by skilled geoscientists.

The discussion of geometric data models often included treatments of the conceptual bases and the theoretical foundations but then also gave, without much differentiation, implementation details [Abel and Smith 1984]. For GIS, two standard models were established: vector and raster methods. Peuquet [1984] even proposed a compromise (vaster) idea. A very extensive literature for efficient implementation of raster models using the quadtree data structure appeared [Samet 1989a; Samet 1989b] and was also applied to 3D problems (octree) [Herbert 1985; Kavouras and Masry 1987].

Efforts to establish a theoretical base for geometric data models started at different points. A landmark work by Corbett [1975] stressed the importance of topology as a basic mathematical concept for organizing geometric data. Another work by Corbett [1979], unfortunately not published in a journal, contains an extensive discussion of implementation at the hardware and assembly language level, which somewhat obscures its deep theoretical contribution. In [Frank 1983] a graph theory concept was found lacking. Peuquet [1988] used image processing concepts and Chan and White [1987] traced the origin of the 'map algebra' concept propagated by Tomlin [1983] back to traditional methods used by urban planners.

The National Center for Geographic Information and Analysis (NCGIA) organized a workshop to discuss the connections between spatial languages and spatial concepts [Mark 1988]. One of the results of this workshop was the recognition of the importance of discussing spatial concepts before any discussion of geometric data models and their implementation. These concepts were further refined in [Mark et al. 1989]. The treatment in this chapter follows and resumes the generic discussion in [Frank and Mark 1990] and applies it to the special case of 3 dimensional geoscientific data.

The problem will first be defined and its complexity detailed in Section 2. In the same section certain underlying assumptions valid for the remainder of the chapter will also be defined and the terminology clarified. In the third section, there will be a discussion of Euclidean geometry, analytical geometry and mathematical topology and how they apply to the organization of geoscientific data. In Section 4, a number of geometric data models using such concepts are formalized. In Section 5, the geometric data structures and their implementation considerations are discussed. The understanding of geometry and the applications to geosciences, especially geology is treated in Section 6. Section 7 concludes the chapter by presenting a discussion on future research needs.

2. Representation of Geometric Data

A representation of the geometric facts geoscientists deal with is needed in the computer system. Computers are essentially machines to execute symbolic computations. A formal system is therefore needed to represent the objects and the relations between them, and rules that can be used to infer other relations between them. This is independent of the programming language one uses or the subject matter that is dealt with. The rules for dealing with the symbols representing the information about reality must be fully determined in order for the computer to apply them. This is much more a problem of conceptualization and structuring of our thinking, than a problem of actually writing a program.

Formalizing a problem to the level that a program can be written is in principle similar to other scientific endeavors: we try to idealize, abstract and formalize with rules some relevant parts of our environment, such that we can deal with it in our thinking. Software engineering is the discipline that has provided us with some guidelines and tools to systematize this process [Sommerville 1989]. Software engineering evolved from the concern with writing small programs - how to instruct a computer to solve a mathematical formulation - so called 'programming in the small'. The most important contribution was probably the concept of structured programming and the ALGOL programming language [Dijkstra 1972; Wirth 1971]. It grew into dealing with the construction of large programs - complexity being the issue - and methods to harness big projects, so called 'programming in the large'. It was observed that writing programs that deal with 'artificial' systems (e.g. banking, insurance, business) is easier than the ones that model 'real', physical systems (e.g. aircraft, GIS, robots). In banking, the computer 'is' the system and its program defines what can be done and how it is done (transactions other than the ones foreseen in the program cannot be carried out). In modeling real systems, one has to assure that the model and the behavior of the real system agree sufficiently to be useful. The program has to foresee all possible reaction - 'not yet programmed' or 'this state not expected' is not an acceptable error message for an aircraft guidance system.

In using artificial intelligence, which is for the purpose considered here just the most advanced form of programming, it was found that problems are difficult to deal with if they [Bobrow, Mittal, and Stefik 1986]:

- need to model spatial aspects
- contain time related reasoning
- require natural language understanding
- deal with human conceptions

Given these precautions, we realize that most of what geology and geosciences in general are interested in is very difficult to model formally. This is caused by the lack of understanding of the relevant base theories (spatial and temporal reasoning), not to speak of a complete formal theory of geology, and the absence of powerful abstractions, as we will see in the following sections.

The problem is thus to understand and formalize the concepts geoscientists use in thinking about the phenomena they are interested in. This leads to the basic question of how people conceptualize the world and the objects in them. Cognitive science and related disciplines have studied these problems and an 'experientalist' point of view is adapted here [Johnson 1987; Lakoff 1987]. In a very simplistic terms, this implies that we study the way people understand the world, and do not concentrate on describing reality per se. Thus the conceptual framework people use to structure their perception of reality becomes part of the object to be studied. Nevertheless the subjective perceptions are comparable and can be effectively communicated, based on the similarity of the fundamental experiences which form the concepts used; for example, all humans observe with eyes similar in physiological construction.

From cognitive science, we gain insight into two unsolved problems, i.e. prototypes and metaphors. Cognitive categories seem to be formed using a prototype or typical exemplar and all other members of the category are more or less similar to this prototype [Lakoff 1987]. This is very different from a set theoretical concept of a category, where all elements share a property and this property constructs the set (e.g. objects having four legs and bark are dogs). It is evident that the cognitive methods to form categories are close to how concepts like animals or mountains are formed. The set theoretic method is the one formalized and used in programming.

The notions established in one context, usually from very concrete experience, can be transferred to another one and used to organize a more abstract situation which in some respects is comparable to the first one [Lakoff and Johnson 1980]. This metaphorical usage of concepts, is reasonably understood in general terms, but it has escaped formalization so far. It is a very important method and often applied to spatial concepts.

Cognitive science has studied the basic relations humans use to structure the world as they observe it. These studies have found that the spatial relations are fundamental. They are mostly based on properties that relate to the human body (inside, in front of, etc.), but they are then used not only to represent the positions of objects with relation to the human body, but for other spatial relations in general. From there, spatial terms are then widely used metaphorically to organize abstract concepts, like social organizations. Thus the analysis of spatial relations is not only relevant for modelling spatial situations but for other situations as well.

The problem thus is to understand the spatial concepts geoscientists use, and to find formal representations, including the spatial and temporal reasoning relevant for them. The problem is difficult to solve as we observe that there is more than a single concept of space and spatial reasoning applicable. We see that human beings use one or the other spatial concept, depending on the circumstances and the problem at hand. In the next section, some of the best known tools to organize spatial knowledge, namely geometry and topology are discussed, and then other concepts that may be more akin to the ones used in 'geologic thinking' are presented.

3. Standard Geometry and Topology

In this section we will concentrate on the geometric aspects and separate them from temporal ones. It is well understood that geology does most of its work in a spacio-temporal continuum and spatial and temporal reasoning are intertwined and linked together. However, each of these topics is a major research problem by itself and there needs to be substantial progress in both before a connection can and should be attempted.

3.1 Euclidian Geometry

Formalizing geometric concepts was attempted by the Greeks, and Euclid gave a complete axiomatization of geometry as it is still taught today in school. Euclid's five axioms are one of the finest examples of abstraction from complex real situations and reduction to a very small number of base concepts, namely points and lines, with only the most relevant properties. It is very obvious that these abstract objects do not exist in reality, but sufficiently good approximations of the behavior of real objects are formalized in the axioms.

Despite the limitations of this axiomatic system and its obvious simplifications, Euclidian geometry appears to be *the* description of geometric reality - despite the fact that there are no infinite straight lines in this world and nobody has ever seen a dimensionless point. During the 19th century, efforts to show the independence of the five axioms resulted in the surprising discovery of 'other' geometries and thus to a deeper discussion of the concept of what constitutes the essence of geometry [Blumenthal 1986] (see Section 6). Despite its mathematical formalization, Euclidian geometry is not very useful for computer implementations. No complete system that allows the automatic proof of all geometric propositions has been established. Its importance for computer implementation is due to the availability of the mapping from geometric to analytical concepts.

3.2 Analytical Geometry

Euclidian geometry discusses the relations of points and lines, and then of simple figures in a plane. These concepts can be mapped to the coordinate plane, which is a mapping from points in the plane to pairs of real numbers. There are analytical solution to all geometric problems, which means that there is a homomorphism between Euclidian objects and operations, and the analytical operations. Thus applying a geometric operation to geometric objects and transforming the product to the analytical form or transforming the original objects and operations to the analytical realm and applying analytical operations to the images result in the same object (Figure 1).

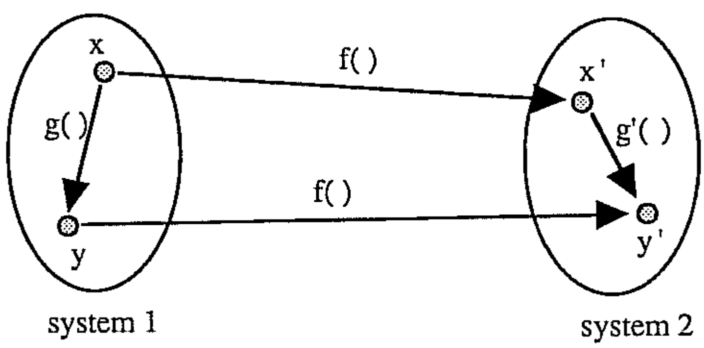

Figure 1: Homomorphism.

Analytical geometry, as founded by Descartes, is usually the basis for the implementation of geometric ideas in a computer system (see section 4.1). It is however important to note that analytical geometry is based on a continuum, i.e. a concept that there is another real number between any two real numbers. This is necessary to fulfil the geometric equivalent, where between any two points is another point (for example the middle point between the original two).

3.3 Topology

3.3.1 Point Set Topology

Another mathematical formalization of spatial concepts is based on the same concept of continuity and infinite sets of points of infinitesimal proximity. Defining sets of such points which form neighborhoods, one can define interesting but simple axiomatizations which represent very different properties from the Euclidian ones, but still geometric in essence.

This so-called point set topology seem not to be very useful by itself, but can be extended to analytical, or combinatorial topology.

3.3.2 Graph Theory

Graph theory deals with points (called nodes) and connections between them (called arcs). A node is said to be incident with an arc. Despite its simplicity, this is a very useful formal set of concepts. It can be used for most navigational tasks (i.e. problems of the type how to get from here to there), as connectedness is the major concept and reasoning about connectedness is simple (connectedness is for example transitive: if we can get from A to B, and from B to C, then we can get from A to C). There is extensive literature about graph theory and its application to practical problems [Deo 1974; Hensley 1973; Johnson 1972].

3.3.3 Combinatorial Topology

Graph theory lacks a concept of space - it just deals with nodes and connections between them. Combinatorial topology uses the base concepts of cells and boundaries between them [Giblin 1977]. These concepts can be used for a representation of space and objects in space with their relations without taking into account the specific position or shape of the objects [Pullar and Egenhofer 1988].

3.4 Taxonomy of Space and 3D Extension

3.4.1 Taxonomy of Geometry

The geometric concepts can be differentiated into:
• topology, based on the neighborhood property and relations between cells, and
• metric, based on a concept of distance between points.
Euclidian geometry is an example of a metric space, as it uses a notion of distance and distance measure between points. A distance is a mapping from two points to a real number such that:

$d(A,A) = 0$ - reflexivity
$d(A,B) = d(B,A)$ - symmetry
$d(A,B) + d(B,C) \geq d(A,C)$ - triangle inequality

Vector spaces, as normally used for the analytical treatment of Euclidian geometry are automatically metric.

3.4.2 Objects vs. Space

Among the different viewpoints humans can use to think about space, two are exemplified in the above sections:
• space as a uniform realm populated by objects, and
• objects which fill space and thus make up 'the space'.
The first one can be associated with a Descartian point of view, the second one with a Kantian. Each of them is valid, consistent and useful, but the two cannot readily be connected.

It should be clear that we deal here only with a concept of objects with sharp boundaries and clear limits. This is the way most geometric reasoning seems to proceed, even understanding that this is not a very realistic assumption, especially for geology where one seems to deal with boundaries of volumes that are not well defined or not precisely known. This is a very important restriction, which needs considerable attention for future research work.

3.4.3 Extension to 3D

The concepts of Euclidean geometry, analytical geometry and topology discussed above were first developed and used widely in 2D applications. The concepts are general enough that they can be extended to 3D cases without difficulty. In Euclidean and analytical geometry, instead of dealing with pairs of real numbers, we now need triplets of real numbers to represent the three axes of 3D space. In topology, we have an additional cell called volume. The treatment however, is much more demanding as there are more special cases to take care of and also visualization in 3D space is limited by the available devices which render projection on 2D surfaces.

4. Formalization of Geometric Concepts

In this section, the formalization of the previously discussed geometric concepts suitable for implementations in computer systems is presented. This process also known as modeling of geometry. Having an axiomatic, mathematical, formal base of the geometric concepts simplifies implementation considerably.

4.1 Finite Computer Systems

There is a major problem related to the assumption of a continuum with an infinite number of points. A computer, as any finite machine, can only represent a finite number of different objects and cannot deal with the infinite number of points potentially necessary for an implementation of analytical geometry. It is useful to remember that the floating point numbers of a computer language are quite different from the real numbers in mathematics. There is a non-countable infinite number of mathematical reals, whereas there is only a finite number of FORTRAN REAL (indeed in 32 bits, there are slightly more integers than floating point numbers).

4.2 Geometric Data Models

In simple terms, the problem of formalizing geometric concepts is how to construct a geometry in a computer that reasonably behave like our notions of geometry. The concept which we call 'geometric data model' is analogous to the concept of data model in database management systems. A data model furnishes the conceptual tools to organize or structure the representation of the application data, e.g. geological data. The geometric data model provides the geometric concepts, which are used to structure the geometric ideas of geology. Unfortunately, the difference between the abstract geometric concepts and the concepts that can be implemented are substantial enough, that we cannot use the abstract geometry (e.g. Euclidian geometry) as a data model; geometry on a integer plane (as opposed to analytical geometry on the plane of real numbers) is quite different, and not yet well understood or formalized in axioms.

There are two major ways to conceptualize a finite geometry: either by dividing the space in a regular fashion or by dividing space in irregularly shaped cells. These two can be linked to the notions previously discussed: the regular division of space stresses the uniform nature of space and thus a Descartian view, and the irregular division of space stresses the object which each occupies a piece of space and thus a Kantian view.

4.2.1 Regular Tessellations

Using regular tessellation to model solid bodies is one method to represent geometric facts in a discrete system. This model is based on the subdivision of the 'universe' of 3D space containing the geo-body into small cells known as volume elements (voxels). Each cell can be either totally inside, partially inside or totally outside of the body. The interior volume is the aggregation of the cells that are totally inside the geo-body. The conceptual smooth boundaries of the body is approximated by staircase boundaries of the cells that are partially inside the geo-body, following the boundaries of the regular cells. This method is significantly different from representing values for a point or an average value for each cell and assume that the values in between are smoothly changing and can be interpolated (see section 4.2.4).

The primitive operations to create, intersect, etc. of cells are provided in the model. Higher level operations on the geo-objects then use these primitive operations.

4.2.2 Irregular Tessellations

This data model is based on the concept of irregular sub-division of space. It is based on the concept of cells as defined by algebraic topology [Alexandroff 1961; Spanier 1966]. Each cell is the simplest polyhedron of a given dimension (Figure 2). These are known as (although terminology varies):

0-cell = point,
1-cell = arc,
2-cell = area,
3-cell = volume.

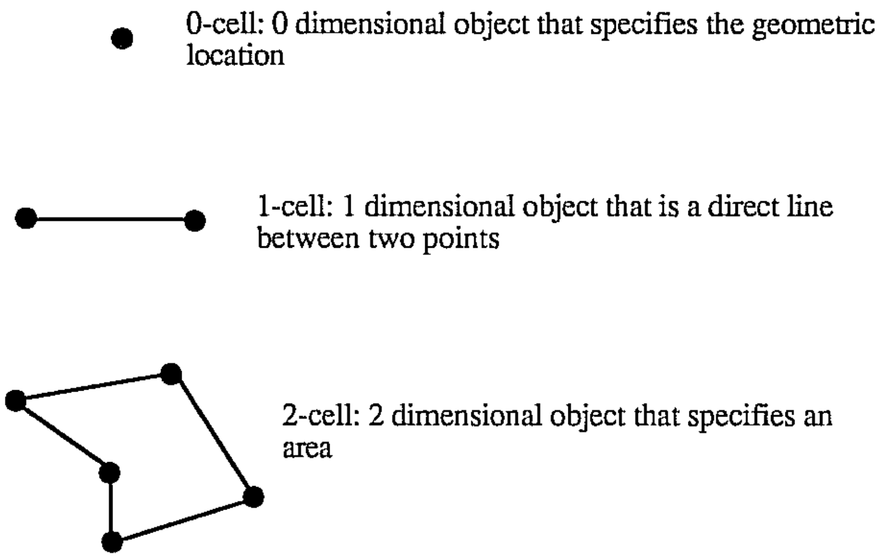

Figure 2: Topological cells.

In principle, the points with coordinate values are used to fix the nodes in space. These nodes are joined by the arcs. The areas are then defined by the arcs that delimit them. References between areas, arcs and nodes allow us to determine the boundary and co-boundary relations. For example, an arc is bounded by a point at each end and an arc is the boundary between two areas. Note that the terminology used in this model is fluid, despite attempts to standardize some terms for the 2D case [USGS 1988]. As in the regular tessellation, the primitive operations on the cells are provided in the model which are then used by the higher level operations.

4.2.3 Spaghetti

This is the simplest data model as it represents only points and lines without connections, and does not have a concept of area or volume. The problem with this model is the inability to deduce other interesting properties, e.g. the intersection of the lines cannot be represented without recourse to the metric (coordinate) representation, which is an approximation; point coordinate values are the result of measuring (which are subject to errors) and computation (which accumulate rounding errors) operations. In this approximation, we can not determine with certainty, if two lines intersect or not [Frank and Kuhn 1986]. This data model is useful to represent the graphical renderings of geoscientific maps, but it cannot help with many of the geometric and analytical operations.

4.2.4 Continuous Function

Phenomena like temperature, gravity field, topography, etc. are thought to be continuous; there is a value at every point in space. The phenomena is represented by a mathematical formula and is often a function of the position in space (x, y, z) and occasionally some other parameters as well. Due to limitations of finite computer systems, this rigorous approach may not be taken, and an approximate method of interpolating the value at the desired point based on some measured discrete values is used.

4.3 Improvements

Moving away from objects we conceive as purely geometric, we can further structure space by allowing some of the following ideas in the geometric data model:

• Aggregates: This refers to the building of hierarchies of objects, such that the aggregate object contains the part objects. The concept is widely discussed in non-spatial context, but we use it here as part of a spatial data model to imply that the aggregate spatially includes the part.

• Errors and uncertainty: Data is known with a certain accuracy and precision. In fact, there are different kinds of uncertainty in spatial data [Robinson and Frank 1985]. The specific problem with geologic data, is that the precision may vary considerably from one data set to another. It is desirable, if not imperative to preserve the precision and some indication for other quality aspects for each data element [Goodchild and Gopal 1989].

• Variable resolution: Resolution contributes enormously to the cost of storing and processing data. Thus representing data of low precision with high resolution is wasteful but the high resolution is needed to deal with the precise data. Thus data structures should be adaptable to the degree of precision of the data set. Quadtree-like structures have a natural concept of resolution built in, which can be exploited to represent data with the resolution needed. Similar ideas have been used for line data [Ballard 1981] and for surface data [Dutton 1988].

• Multiple representations: The same object may be represented in various ways, using the same or a different data model, depending on the task we intend to accomplish. It may be appropriate to have more than one representation of the same object in a single system and connect these multiple representations such that for each task the most appropriate one is selected and that changes in one propagates into the other ones [Bruegger and Frank 1989]. This would lead to the notion that a system can contain more than one geometric concept (and in consequence more than one implementation). Indeed in the 2D GIS, the unification of regular and irregular tessellation-based concepts, under the heading of integration of raster and vector systems, is one of the most important areas of research [Aronoff, Mosher, and Maher 1987; Ehlers, Edwards, and Bedard 1989]

5. Geometric Data Structures for Implementations

Geometric data structures are the specific implementations which provide the operation demanded in the geometric data model using specific storage structures and algorithms. From a conceptual point of view, all geometric data structures that implement the same data model should be equal - they should offer the same operations and should compute the same results. They may however differ in the amount of resources they utilize or the speed of execution. In this section, we discuss some alternatives for geometric data structures. Geoscientists should not concentrate on these aspects, but may want to understand the methods employed. In general, computer scientists are working on such problems and need from the application domain specialist, i.e. the geologists, only formal descriptions of the problems requiring a solution [Buchmann et al. 1990].

5.1 Regular Tessellation

In 2D cases, a number of different regular tessellations have been studied [Diaz and Bell 1986], but very few systems use other than the square regular cell structure. In 3D applications, only cube structures (the 3D equivalent of squares) have been widely studied and used.

The major consideration for implementation is the use of the redundancy in the data to reduce data storage and processing, without loosing resolution. This is even more important in 3D than in 2D applications, as the amount of data increases even more rapidly. To represent an area of 100x100 km and to a depth of 5000 m with a resolution of 1000 m cubes, we need $50x10^3$ cells; if we increase the resolution to 100m, we create $50x10^6$ cells and if we need a 10m resolution, then we will have to deal with $50x10^9$ cells. It is important to note that the Nyquist law applies here as well: no detail smaller than the size of two cells can be included and we will have to filter with a high pass everything of higher spatial frequency from the data. Thus finding the appropriate balance between resolution and storage and processing requirements is a difficult task. The processing of the data, for example in the 'overlay' operations is essentially linear in the amount of data stored; thus increasing resolution does not only increase the storage requirements, but also the processing time.

In a regular tessellation, we record the values for the properties we are interested in for each cell. This contains considerable redundancies, as the auto-correlation from cell to cell is very high; most of the time the neighbors of a cell contain the same or very similar values. In order to exploit this characteristic, a method of run length encoding can be extended to a 3D structure: the cells are traversed in a row by row fashion and a sequence of values for a property generated. This sequence will contain subsequences of the same value - these subsequences are then replaced by a count and a single value. Algorithm to combine two structures encoded in this form can be easily devised.

Another encoding method to capture large areas with similar values is based on quadtree and its extension to 3D octree. The data structure is generated by the following recursive procedure:

```
procedure decompose(volume);
if uniform property value for volume
  encode volume property
else
  begin
    split volume into octants;
    for each octant
    decompose(octant);
  end;
end decompose;
```

From a point of view of generalization, we can see that these methods are all variants on the theme of:
1. Find a method to traverse the cells (e.g. row by row or in a 'Morton' sequence for the octree).
2. Identify sequences of similar values and replace them with a shorter mark, which encodes the length of the run.
3. Encode the new sequence (e.g. as pointer octree, linear octree etc.).

The major criticism of these methods is their storage requirements. Despite considerable compression, data sets of a realistic size for geological applications do become very large. The octree is also not very suitable for fragmented ore body due to its deficiency in modeling the boundary representation; being not invariant to translation and rotation (the method uses approximations when converting to a boundary representation) made octree implementation not suitable for surface analysis [Kavouras and Masry 1987].

5.2 Irregular Tessellation

In 2D, this is a method widely used with large number of commercially available implementations. In a GIS, it is usually called a topological data structure, because it uses topology and the 'boundary' and 'co-boundary' relations.

Two major variations exist in the present implementations of topological data structures. First is the variation in the exact format that the boundary/co-boundary references are computed, i.e. which ones are stored and which one is computed. The second variation is the limitations and restrictions imposed from the most general cell model; while some allows polygons of any dimension to exist, some require that objects must be triangulated. The reasons for these differences lie in the problems of implementing this model, especially dealing with the special cases.

We can allow or disallow curved lines between two nodes (or curved surfaces in the case of 3D). This is not a limitation in concept but in the data model, because we can only approximate curved lines and surfaces with an appropriate number of additional points.

The base operation in this structure is the intersection of cells. The problem is that, intersection between whatever forms we allow must be computable and representable in the same model:
• representable - in order for the model to be closed under intersection operation, the result of the intersection of two object must have a boundary that is again representable in the model. For example, a model must have provision for ellipse when allowing intersecting a cylinder with an inclined plane.
• computable - the result of intersection must always be computable. This is not even true for the intersection of two straight lines [Nievergelt and Schorn 1988]. Solutions require some limitation in the original data, e.g. we can ask that two points are always sufficiently

separated that an intermediate point can be computed. This is obviously limited to a single intersection and must be tested again for the next intersection.

Conceptually allowing volumes of arbitrary shape is the most desirable as it imposes the least restrictions on modelling; unfortunately it is impossible to implement in a pure form. Simplifications are possible in different directions:

- approximation of arcs by straight lines and areas by planes or by class of curves. Splines especially NURBS, are currently a popular choice. Important is that the class is closed under intersection.
- the most general cell model leads to a large number of special cases, each of which must be dealt with (e.g. a volume may contain a hole, which touches the exterior surface in a point or an edge). This makes coding tricky and execution is slowed by testing for these cases. One may restrict the complexity by disallowing holes or volumes with overly complex shapes and forcing that objects are broken in similar parts. The extreme of this is a triangulation model, where all volumes are constructed from tetrahedrons (the simplest volume) [Frank and Kuhn 1986]. This model sidesteps all the special cases and can be implemented dimension independent - the same code works for 2D or 3D and does not become more difficult for volume treatment. It can be based on a very elegant algebra, the 'chains', which is in its mechanics very similar to the manipulation of polynoms in regular algebra. The cost for this simplicity is the increase in volume elements one has to treat - a complex volume has to be divided into a number of tetrahedrons, which are then treated individually.

Irregular tessellation models can be combined with concepts to represent non-manifold geometry, which appear to be important for geological modelling.

6. Geometry in Geosciences

6.1 What is Geometry

It has been shown that there is more than one concept used to represent geometric facts and geometric reasoning. We should now ask, what is the essence of geometry and how is it used in geosciences. The different geometric data models represent different aspects and depending on what aspects we model, one or the other problem becomes more easy to resolve. We may follow the example of modern physics, where it was shown how properties can be declared 'geometric' (e.g. the distribution of mass) in order to make formulae that describe other non-geometric aspects simpler (Lorentz's transforms).

In the last century, mathematicians discovered that there is not only a single geometry (Euclidean) but many different ones that can be constructed. It then became an interesting problem to identify what makes a formal system become a geometry. Since the work of Felix Klein [Klein 1872], geometry is predominantly seen as the study of invariant under certain transformations [Blumenthal and Menger 1970]. The idea of transformation is deeply rooted in the human conception of space, as the process of measuring distances, areas and volumes is seen to rely on the comparison of unknown quantities with measuring tools which are considered to have invariant size. Euclid's congruence theorems constitute an early incorporation of invariance into geometry.

6.2 Geometric Concepts in Geosciences

At this point it seems appropriate to discuss the geometry we encounter in geoscientific applications: what are the geometric concepts geoscientists use? It is obviously difficult to do, for both the outsiders and the geoscientists themselves. For the geoscientists, these concepts are so familiar that they do not think of them in isolation, and for the outsiders, as they observe the use of geometric objects, they may not understand the finer points of their use.

The problem and the approach is similar to the discussion in [Frank, Palmer, and Robinson 1986], where the formalization of the spatial concepts useful for geomorphology was attempted. The stress in that paper was to show, that methods can be formalized and formal reasoning applied. It was argued that from such an approach, rigor in the analytical work can be gained and the results made easier to compare, containing less subjective interpretation. In geomorphology, a formal concept for the representation of the surface of the earth is necessary. A triangulated surface representation was proposed, as this made some definitions more intuitive, but a raster surface could have been used as well - with differences in the definitions and the results. Remember that the formalization influences the structuring and thus the way we conceptualize reality as we perceive it.

Geological work - from a layman's perspective - deals with processes in space and time. The problem seems to be the construction of a plausible sequence of processes that would result in a situation which corresponds to the observable facts. Note that geological observations are very few and usually widely spread in space (and only a single point in time). It has been said, that the geologic graphical products (e.g. geological maps and diagrams) are difficult to read because they truly represent a sequence of processes but then show only the last stage - and it remains to the viewer, to unravel the past [Simmons 1982].

The currently available applications of GIS technology to geology is similarly limited. It becomes feasible to model in 3D and make these models visible in various ways, but we cannot include the process that created the current situation in the model. Current GIS technology does not include models of changes or processes nor do they include a notion of time. Research in this direction is planned [NCGIA 1989].

In general, the formal descriptions of geological processes do not allow for inversion in the time domain, i.e. we can not postulate a current situation and then solve for the previous state. It is the geologists intuition and training that determines which sequence of processes could have happened and what starting situation was appropriate. In a geological GIS, the automatic inversion in time of geological processes is not necessary, but we need models of processes such that a geologist can interpret a situation and determine which sequence of changes could have created it. This sequence can then be played forward and backward and applied uniformly.

In an actual system [Pflug 1988], the concept of layers, overlaying and folding are included. The modelling starts with a base layer and some assumptions of layer thickness. This is then intersected with the observed terrain surface and produces a series of geological maps; perspective views, isometric views of orthogonal sections and the individual section. The model can be compared with the observed outcrops and duly adjusted.

6.3 Data Models Revisited

Data models are more or less appropriate for geological modelling, depending on how well they capture the essential aspects of geological spatial thinking. It is not necessarily the most general model that is best. Two spatial concepts geologists seem to use often, namely (1) boundary surfaces for volumes or discontinuity surfaces, and (2) layers of more or less homogeneous material, can be used to construct a 3D models for geology.

6.3.1 $2\frac{1}{2}$ D Data Models

The previous discussion assumed a homogeneous and isotropic space; a space that has the same properties in all directions and that is treated equally for all directions (or at least close, as a regular tessellation has some preferred directions, but conceptually it tries to mimic an isotropic space).

The object of study in geology, a thin layer on the outer rim of the earth is in most instances more extended in 2 dimensions (x, y) then in the third one (z). It is therefore often attempted to use a different data model for the 2 major dimensions than for the third one. This is often called 2.5 D models and has nothing to do with Hausdorff or fractal dimension, where 2.5 would have a different signification. Typically such systems represent a 3D space with 2D surfaces, adding a third dimension to each node, and agree on some interpolation method in between these points.

6.3.2 Layers

We could assume that the objects we deal with are layers which are relatively thin compared to their horizontal extension and lay essentially flat. They lay one atop the other and the boundary of one is also the boundary of the other. We could model these boundary surfaces individually, but the model would then not contain an explicit representation of volumes. To model volume but not going to a full isotropic 3D volume representation, we may exploit the anisotropy in the situation. If the geometric data model is restricted to triangulation in the plane, we can use prismatic volume elements, allowing arbitrary cells in the horizontal plane and vertical limits for the prism (Figure 3).

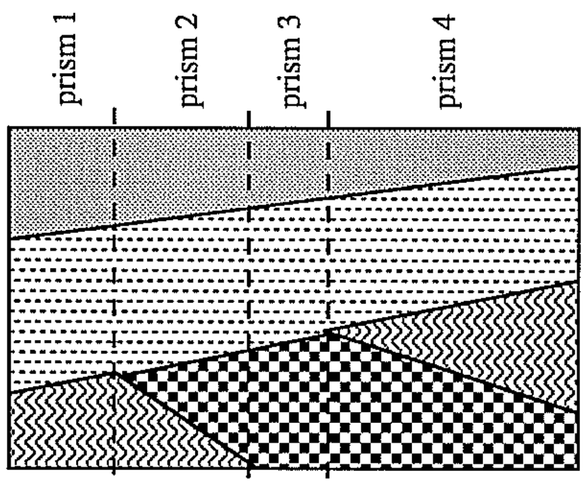

Figure 3: Prismatic volume concept.

The prisms are each formed by projecting all layer boundary intersections to the ground. On these areas of constant sequences of layers one erects a vertical prism. This is simple to implement as there is only one geometric layer (in the horizontal plane) and z values to define the intersection of the prism with the boundaries of the layers. The problem is how to deal with deviations from the basic assumption in the model, namely layers which end, and layers which fold over or fall in vertically. We may form aggregates from disjoint parts of a layer [Carlson 1987].

7. Future Research Needs and Conclusions

7.1 Geometric Modelling

Geologist often assume that the volumes they deal have well defined boundaries. This is not only a convenience for their graphical representation but must also influence the way geologists reason about them. One may build modelling systems, that include this ideal property. On the other hand, geologic data is deduced from very few point samples and other indices. The location of boundaries is interpolated between sparse points and often not precisely known. A geological GIS could be built to allow uncertainty in the location of boundaries and use this uncertainty to reconcile contradictory evidence to deduce better data. Thirdly, the assumption of well defined boundaries for geo-bodies is not always justified. Many phenomena geologist deal with are changing gradually in space and do not have sharp boundaries. There is a need for modelling such situations.

Of course, some may wish to combine these three 'view points' in a single system, which allows movement from one to the other freely. This may be a long term goal, but not necessarily achievable today. First we have to understand and build individual systems, each dealing with one of these aspects. Then we may be able to link them, such that representations of a single object in one or the other are tied together and properties may be transferred from one to the other. Ultimately one might construct a conceptual shell that englobes all three 'view points' in an uniform manner. The difficulty is not only in the programming but also in constructing interfaces that are useful for geologists.

7.2 Process Models

A second line of extension of current 3D modelling tools is toward modelling of processes important in geology. A static 3 dimensional view is helpful for geologist, but it appears that most reasoning is in terms of processes, that change geometry.

One first thinks about elaborate tools that deal with folding of layers in full 3D, preferably with attractive graphical presentations. This may be asking for much - given the problems with representation of geometry discussed above - but simpler, less elaborate tools may be feasible. If we assume that much of geological thinking is not dependent on the exact location or extension of the bodies - and given our limited knowledge of their boundaries, it cannot - but just on some of their special properties. Obvious examples are sequences of layers, where the order of the layers is predominant and the layer thickness etc. is of less importance and much more variable. This is in general terms a 'topological relation', a relation of sharing a boundary, and not dependent on exact measurements. Discrete mathematics and especially symbolic reasoning can be used on such data to deduce qualitative (not quantitative) results. Symbolic reasoning could be extended to include qualitative formulation of processes on a level similar to the simplified block diagrams included in geological texts.

On the quantitative and geometric side, one could include other properties, which must remain invariant into the geometry. Obvious candidates seem to be constancy of volume or mass, gravity as a dominant force, etc.

Crucial to all formulation of process is modelling of time in a GIS. Requirements from geology are certainly different then say the requirements from the legal side to model 'property ownership' in time. Further study is necessary to determine which of the different models is most appropriate [Barrera and Al-Taha 1990].

7.3 Final Remarks

Geology poses very interesting demands of a GIS: modelling of geometry with uncertainty and vagueness in the boundaries, modelling of time and processes etc. However, further understanding of each of these aspects is required before integration of them is attempted. On the other hand, results from 2D GIS and CAD/CAM and the systems that have been constructed, are useful for geological applications.

Acknowledgements

The authors would like to thank the National Science Foundation for financial support under NCGIA grant No. SES-8810917 and additional support from the Intergraph Corporation.

References

Abel, D., and J. Smith. (1984). "A Data Structure and Query Algorithm Based on a Linear Key for a Database of Areal Entities." *The Australian Computer Journal*, vol. 16, no. 4, pp. 147-154.

Alexandroff, P. (1961). *Elementary Concepts of Topology*. New York: Dover Publications, Inc.

Ansaldi, S. et al. (1985). "Geometric Modelling of Solid Objects by Using a Face Adjacency Graph Representation." *Computer Graphics*, vol. 19, no. 3, pp. 131-139.

Aronoff, S. et al. (1987). "Operational Data Integration - Image Processing to Interface Vector GIS and Remotely Sensed Data." *Proceedings of the GIS '87*, vol. 1, pp. 216-225, San Francisco, CA, Oct 26-30.

Ballard, D. H. (1981). "Strip Trees: A Hierarchical Representation for Curves." *Communication of the ACM*, vol. 24, no. 5, pp. 310-321.

Barrera, R., and K. Al-Taha. (1990). "Models in Temporal Knowledge Representation and Temporal DBMS." *Report of the ASDS Project to Digital Equipment Corporation*. Department of Surveying Engineering, University of Maine, Orono, ME.

Blumenthal, L. M. (1986). *A Modern View of Geometry*. New York: Dover Publications, Inc.

Blumenthal, L. M, and K. Menger. (1970). *Studies in Geometry*. San Francisco: Freeman and Company.

Bobrow, D. G. et al. (1986). "Expert Systems: Perils and Promise." *Communications of the ACM*, vol. 29, no. 9, pp. 880-894.

Bruegger, B. P., and A. U. Frank. (1989). "Hierarchies over Topological Data Structures." *Proceedings of the ASPRS-ACSM Annual Convention,* vol. 4, pp. 137-145, Baltimore, MD, April 2-7.

Buchmann, A. et al. Ed. (1990). *Design and Implementation of Large Spatial Databases.* New York: Springer-Verlag.

Carlson, E. (1987). "Three Dimensional Conceptual Modeling of Subsurface Structures." *Proceedings of the ASPRS-ACSM Annual Convention,* vol. 3, pp. 188-200, Baltimore, MD, Mar 29 - Apr 3.

Chan, K. K. L., and D. White. (1987). "Map Algebra: An Object Oriented Implementation." *Proceedings of the IGIS: The Research Agenda,* vol. 2, pp. 127-150, Arlington, VA, Nov 15-18.

Corbett, J. (1975). "Topological Principles in Cartography." *Proceedings of the 2nd International Symposium on Computer-Assisted Cartography,* pp. 61-65, Reston, VA, Sep 21-25.

Corbett, J. (1979). "Topological Principles of Cartography." Technical Report 48, Bureau of Census, Department of Commerce, Washington DC.

Deo, N. (1974). *Graph Theory with Applications to Engineering and Computer Science.* Englewood Cliffs: Prentice Hall.

Diaz, B. M., and S. B. M. Bell. Ed. (1986). *Spatial Data Processing using Tesseral Methods (collected papers from Tesseral Workshops 1 and 2).* Reading, UK: Natural Environment Research Council, Unit for Thematic Information Systems.

Dijkstra, E. W. (1972). "Notes on Structured Programming." In *Structured Programming.* Edited by O. J. Dahl, et al. New York: Academic Press.

Driel, N. V. (1975). "Geological Information in Computer Mapping System for Land Use Planning." *Proceedings of the 2nd International Symposium on Computer-Assisted Cartography,* pp. 491-495, Reston, VA, Sep 21-25.

Dutton, G. Ed. (1978). *First International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems.* Cambridge: Harvard University.

Dutton, G. (1988). "Modelling Locational Uncertainty via Hierarchical Tessellation." Paper Read at the Specialist Meeting of Initiative 1, NCGIA: Accuracy of Spatial Databases, Montecito, CA, Dec 12-16.

Ehlers, M. et al. (1989). "Integration of Remote Sensing with Geographic Information Systems: A Necessary Evolution." *Photogrammetric Engineering and Remote Sensing,* vol. 55, no. 11, pp. 1619-1627.

Frank, A. U. (1983). "Data Structures for Land Information Systems: Semantical, Topological and Spatial Relations in Data of Geo-Sciences" (in German). Ph.D. Thesis, Swiss Federal Institute of Technology, Zurich.

Frank, A. U. (1984). "Computer Assisted Cartography: Graphics or Geometry?" *ASCE Journal of Surveying Engineering,* vol. 110, no. 2, pp. 159-168.

Frank, A. U., and W. Kuhn. (1986). "Cell Graph: A Provable Correct Method for the Storage of Geometry." *Proceedings of the Second International Symposium on Spatial Data Handling*, pp. 411-436, Seattle, WA,

Frank, A. U., and D. M. Mark. (1990). "Language Issues for Geographical Information Systems." In *Geographic Information Systems: Principles and Applications (in press)*. Edited by Maguire, et al. New York: Longman.

Frank, A. U. et al. (1986). "Formal Methods for Accurate Definition of Some Fundamental Terms in Physical Geography." *Proceedings of the Second International Symposium on Spatial Data Handling*, pp. 583-599, Seattle, WA, Jul 5-10.

Gargantini, I. (1982). "Linear Octrees for Fast Processing of Three-Dimensional Objects." *Computer Graphics and Image Processing*, vol. 20, no. 4, pp. 365-374.

Giblin, P. (1977). *Graphs, Surfaces and Homology*. London: Chapman and Hall.

Goodchild, M. F., and S. Gopal. Ed. (1989). *Accuracy of Spatial Databases*. Bristol: Taylor & Francis.

Hensley, E. J. (1973). *Graph Theory in Modern Engineering; Computer Aided Design, Control, Optimization, Reliability Analysis*. New York: Academic Press.

Herbert, F. (1985). "Solid Modeling for Architectural Design using Ocpaths." *Computers & Graphics*, vol. 9, no. 2, pp. 107-116.

Johnson, D. E. (1972). *Graph Theory with Engineering Applications*. New York: Ronald Press.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.

Kavouras, M., and S. Masry. (1987). "An Information System for Geosciences: Design Considerations." *Proceedings of the 8th International Symposium on Computer-Assisted Cartography*, pp. 336-345, Baltimore, MD, Mar 29 - Apr 3.

Klein, F. (1872). *Comparative Considerations about Recent Geometric Investigations (in German)*. Enlargen, F. R. Germany:

Lakoff, G. (1987). "Cognitive Models and Prototype Theory." In *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Edited by U. Neisser, pp. 63-100. New York: Cambridge University Press.

Lakoff, G., and M. Johnson. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Mark, D. M. (1988). "Cognitive and Linguistic Aspects of Geographic Space." Report on a Workshop of the National Center for Geographic Information and Analysis, SUNY at Buffalo.

Mark, D. M. et al. (1989). "Languages of Spatial Relations: Initiative Two Specialist Meeting Report." *Technical Report 89-2*. National Center for Geographic Information and Analysis, SUNY at Buffalo.

Meagher, D. (1982). "Geometric Modelling Using Octree Encoding." *Computer Graphics and Image Processing*, vol. 19, no. 2, pp. 129-147.

NCGIA. (1989). "The Research Plan of the National Center for Geographic Information and Analysis." *International Journal of Geographic Information Systems*, vol. 3, no. 2, pp. 117-136.

Nievergelt, J., and P. Schorn. (1988). "Line Problems with Supra-Linear Growth" (in German). *Informatik Spektrum*, vol. 11, no. 4, pp. 214-217.

Peuquet, D. (1984). "A Conceptual Framework and Comparison of Spatial Data Models." *Cartographica*, vol. 21, pp. 66-113.

Peuquet, D. (1988). "Toward the Definition and Use of Complex Spatial Relationships." *Proceedings of the Third International Symposium on Spatial Data Handling*, pp. 211-223, Sydney, Australia, Aug 17-19.

Pflug, R. (1988). "Solid Modeling of Geological Objects with 3D Rasters." *Geologische Jahrbuch*, vol. 104, pp. 213-219.

Pullar, D., and M. J. Egenhofer. (1988). "Towards Formal Definitions of Topological Relations Among Spatial Objects." *Proceedings of the Third International Symposium on Spatial Data Handling*, pp. 225-242, Sydney, Australia, Aug 17-19.

Requicha, A. (1980). "Representation for Rigid Solids: Theory, Methods and Systems." *ACM Computing Surveys*, vol. 12, no. 4, pp. 437-464.

Robinson, V., and A. U. Frank. (1985). "About Different Kinds of Uncertainty in Collections of Spatial Data." *Proceedings of the 7th International Symposium on Computer-Assisted Cartography*, pp. 440-449, Washington, DC, Mar 11-14.

Samet, H. (1989a). *Applications of Spatial Data Structures: Computer Graphics, Image Processing and GIS*. Reading: Addison-Wesley.

Samet, H. (1989b). *The Design and Analysis of Spatial Data Structures*. Reading: Addison-Wesley.

Simmons, R. G. (1982). "Spatial and Temporal Reasoning in Geologic Map Interpretation." *Proceedings of the National Conference on Artificial Intelligence*, pp. 152-154, Pittsburg, PA, Aug 18-20.

Sommerville, I. (1989). *Software Engineering*. New York: Addison-Wesley.

Spanier, E. (1966). *Algebraic Topology*. New York: McGraw-Hill.

Thomas, A. L. (1978). "Data Structures for Modelling Polygonal and Polyhedral Objects." In *First International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems*. Edited by G. Dutton. Cambridge: Harvard University.

Tipper, J. C. (1979). "Surface Modelling Techniques." *Series on Spatial Analysis, Vol. 4*. Kansas Geological Survey, Lawrence.

Tomlin, C. D. (1983). "Digital Cartographic Modeling Techniques in Environmental Planning." Ph.D. Thesis, Yale University, New Haven.

USGS. (1988). "The Proposed Standard for Digital Cartographic Data." *The American Cartographer,* vol. 15, no. 1, pp. 21-31.

Weiler, L. (1985). "Edge Based Data Structures for Solid Modeling in a Curved Surface Environment." *IEEE Computer Graphics and Applications,* vol. 5, no. 1, pp. 21-40.

Weller, B. (1975). "Computer-Assisted Information System and Computer-Assisted Cartography: Tools or Tinker Toys of Urban Governance." *Proceedings of the 2nd International Symposium on Computer-Assisted Cartography,* pp. 482-490, Reston, VA, Sep 21-25.

Wirth, N. (1971). "The Programming Language Pascal." *Acta Informatica,* vol. 1, no. 1, pp. 35-63.