## Spatial Data models for GIS: Representation in computer

When Geographical data are entered into a computer the user will be most at ease if the geographical information system can accept the phenomenological data structures that he has always been accustomed to using. But computers are not organized like human minds and must be programmed to represent phenomenological structures appropriately. Moreover, the way the geographical data are visualized by the user is frequently, not the most efficient way to structure the computer database. Finally, the data have to be written and stored on magnetic devices that need to be addressed in a specific way.
Geographical information systems provide methods for representing spatial data that allow the user to adopt conceptual models resembling to a large extent the three classes of model as discussed above. There are two broad categories of spatial data models. These are vector data model and raster data models.
The data base concept is central to a GIS and is the main difference between a GIS and drafting or computer mapping systems, which can produce only good graphic output. All contemporary Geographic Information system incorporates a data base management system. Data base systems provide the means of storing a wide range of geographic information and Updating it without the need to rewrite program. In GIS, the spatial data models handle where the features are and Nonspatial data models or Data base management system handle the feature description and how each feature is related to other.
Two approaches or models have been widely adopted for representing the spatial data within GIS ; The cartographic map model and the geo-relational model. Each of these approaches is based on a specific spatial data model. The Composite Map Model is usually based on a tessellated (raster) representation of space and the Geo-relational model is usually associated with a vector representation of space.
Vector data model represents phenomena in terms of the spatial primitives, or components, consisting of point, line, polygon, surfaces and volumes. Raster data model represents phenomena as occupying the cells of a predefined, grid shaped tessellation.

## Raster Spatial Data Model

Typically the grid-cell tessellation, widely known as raster structure is the most commonly adopted structure in a GIS package. The quadtree tessellation is another method, which has been adopted by many GIS packages.
Raster based spatial models regard space as a tessellation of cells, each of which is associated with a record of the classification or identity of the phenomena that occupies it. The raster model represents the two- dimensional location of phenomena as a matrix of grid cells. Each cell stores a data item defining the entity, the class or the value of the represented phenomena. Each cell is also known as a ‘pixel’.
Raster models also use layered approach. In each layer raster cells represent

the presence or absence of phenomena of particular class. The value of particular cell indicates the categories of phenomena within the given class.
Since the cells are of a fixed size and location, rasters tend to represent natural and human made phenomena in a blocky fashion. The extent to which the distribution of phenomena may be mis-represented depends upon the size of the cells relative to the feature of interest. If the cells are sufficiently small relative to the feature, the raster may be a particularly effective means of representing the often somewhat random distribution of the boundaries of natural phenomena that may tend to merge gradually into each other, rather than being neatly delineated.
The Grid-cell or raster model is a relatively simple approach to data representation, both conceptually and operationally, and it has therefore, been popular since the earliest days of GIS development. The model is currently implemented in a large number of raster-based GIS packages. Raster data performs a discretisation of the geometric area of interest and the entire space is broken into grid cells of a fixed or uniform size. In this type of representation of geographic data, a set of cells located by coordinates is used, each cell is independently addressed with the value of an attribute by specifying values to each grid cell.

The simplest raster data structures consist of an array of grid cells. Each grid cell is referenced by a row and column number and it contains a number representing the type or value of the attribute being mapped. In raster structures a point is represented by a single grid cell; a line by a number of neighbouring cells strung out in given direction and an area by an agglomeration of neighbouring cells.

Rasters are limited by the area they can represent and also the limits of storage space. Also the fineness of data is limited by the cell size, thus the area of coverage is traded off with the resolution of the coverage. The storage problems are handled by resorting to coding, such as run-length ending, chain coding, block coding etc.
The capabilities of a raster-based Cartographic modeling system ultimately arise from functions associated with individual data-transforming operations and the way in which these operations are combined. This transformation of data is facilitated by the fact that map layer zones are represented not by lines or symbols, but by numerical values. It is also facilitated by the fact that these values are directly associated with individual locations. The use of numbers here makes it possible to transform geographical characteristics using mathematical and arithmetical function. Also, it is very easy to carry out overlay operations to compare attributes recorded in different layers. Each attribute associated with the grid cell can be combined logically or arithmetically with attributes in corresponding cells of the other layers to create a new attribute value for the resulting overlay.

The limitations of raster approach lie in that, if each cell is confined to a single classification, the raster model may still fail to represent adequately the transitional nature of change of some natural phenomena. Unless sampling is reduced to a microscopic level, many classes of data, such as soils, sediments and vegetation, are in fact mixtures of categories. Such fuzzy characteristics can be represented more effectively in a raster by means of mixed pixels, in which the component categories are represented by measured or expected percentages of the total composition of the cell.

**Vector Spatial Data model**

In the second and more intensely developed approach to information integration, attribute information is associated with point, line and polygons – as spatial entities that describe features occurring in the real world. Thus, for example, a point feature such as a city may have associated with it such items of information as its total population, number of houses, number of schools and so on. Similarly, a linear feature such as a river might have associated with it such information as name, mean discharge etc. A polygonal feature such as landuse category might be linked to information describing its use, past land use, its soil type and so forth.
The vector representation of an object is an attempt to represent the object as exactly as possible. The co-ordinate space is assumed to be continuous, not quantised as with the raster space, allowing all positions, lengths, and dimensions to be defined precisely.
Vector data models treat phenomena as sets of primitives and compose spatial entities. In 2D models the primitive entities are points, lines and areas. The vector data structure represents each geographical feature by a set of coordinates. Vectors as x,y coordinates define points, lines and polygons. The basic premise of the vector based structuring is to define a 2-dimensional space where coordinates on the two-axes represent features. Generally, representing points and lines is straightforward – points are characterized by a x,y coordinate pair and line by a set of x,y coordinate pairs with a specific beginning and ending vector. However, representing polygons in vector storage poses a challenge. The three vector structures used in Geographical Information Systems (GIS) for the storage of points. lines, and areas are:

**Point entities**
Point entities can be considered to embrace all geographical and graphical entities that are positioned by a single XY co-ordinate pair. Besides the XY coordinates, other data must be stored to indicate what kind of 'point' it is, and the other information associated with it (figure 8)

**Line entities**
Line entities can be defined as all linear features built up of straight line segments made up of two or more co-ordinates. The simplest line required the

storage of a begin point and an end point (two XY coordinate pairs) plus a possible record indicating the display symbol to be used.

An **'arc'**, a **'chain**' or a '**string'** is a set of n XY co-ordinates pairs describing a continuous complex line. the shorter the line segments, and the larger the number of XY co-ordinates pairs, the closer the chain will approximate a complex curve. Data storage space can be saved at the expense of processing time by storing a number that indicates that the display driver routine should fit a mathematical interpolation function to the stored co-ordinates when the line data are sent to the display device.

As with 'point' and simple line, chains can be stored with data records indicating the type of display line symbols to be used.

There are several ways of vector structures possible for structuring polygons.

The simplest way to represent a polygon is the **spagetti representation,** which is nothing, but an extension of the simple chain, i.e. to represent each polygon as a set of XY co-ordinates on the boundary i.e. the polygons are discretised to the concept of line representation and are characterized by a set of xy coordinate pairs, but have the same vector as the beginning and ending vector that is representing a self closing line as a polygon. The name of the symbols are used to tell the user what each polygon in ard then held as a set of simple text entities (refer figure **7**).

**Topology**

An important aspect of vector-based models is that they enable individual components to be isolated for the purpose of carrying out measurements of, for example, area and length, and for determining the spatial relationships between the components. Spatial relationships of connectivity and adjacency are examples of topological relationships and a GIS spatial model in which these relationships are explicitly recorded is described as topologically structured. In a fully topologically structured data set, wherever lines or areas cross each other, nodes will be created at the intersections and new areal subdivisions defined. In two dimensions, this may be regarded as part of the process of planar enforcement referred to previously.

Vector-based spatial data models that are topologically structured are often described in a terminology of topological objects or primitives, which are classified in terms of topological dimensions, where an n-cell topological object is of ndimensionality. Two-dimensional topological objects consist of polygons (faces or areas), arcs and nodes. For areal information, adjacency between the areas can be recorded in terms of feature codes associated with the left and right sides of arcs. The expressions 'left' and 'right' are given meaning in this context by specifying the direction of the arc in terms of a 'from node' and a 'to node'.

The composition of each polygon can be defined by listing the component arcs, including a negative sign where necessary to ensure consistency in arc direction. In order to distinguish between external and internal boundaries, a convention of clockwise for the former and anticlockwise for the latter (or vice versa), can be adopted. For the purpose of network analysis, each node may be associated with a list of the arcs, which it bounds. The list of arcs connected to each node will generally be in a predetermined order, namely clockwise or anticlockwise.

Topological structure is important in keeping track of the components of complex objects and in determining the spatial relationships of connectivity and adjacency between recorded phenomena. Thus if two lines cross each other they will share a common node. If two areas are adjacent to each other, such as two neighboring counties, they will share a common boundary arc. If the boundary of a county coincides with the path of a river they might also share the same arc. The inclusion of one area in another, such as a specific type of forest within a county, will result in their sharing common polygons.

The presence of these various spatial relationships can be determined by relatively simple comparisons of the identifiers of their topological components, rather than requiring possibly computationally demanding geometric calculations based on coordinates. It may also be noted that because shared spatial objects are only stored once, though perhaps referenced many times, storage space is saved by avoiding duplication of the same geometric data. This in turn assists in the maintenance of the integrity of the database by avoiding the possibility of two different versions of the same geometric components.

**The choice between Raster and Vector**

The raster and vector methods for spatial data structures are distinctly different approaches to modeling geographical information, but are they mutually exclusive? Only a few years ago, the conventional wisdom was that raster and vector data structures were irreconcilable alternatives. They were then irreconcilable because raster methods were required huge computer memories to store and process image at the level of spatial resolution obtained by vector structures. Certain kinds of data manipulation, such as polygon intersection or spatial averaging presented enormous technical problems with the choice of raster methods that allowed easy spatial analysis but resulted in ugly maps, or vector methods that could provide database of manageable size and elegant graphics but in which spatial analysis was extremely difficult.

**Vector methods**

**Advantages :-**

- Good representation of phenomenological data structure
- Compact data structure
- Topology can be completely described with network linkages
- Accurate graphics
- Retrieval, updating and generalization of graphics and attributes are possible

**Disadvantages:**

- Complex data structures
- Combination of several vector polygon maps or polygon and raster maps through overlay creates difficulties
- Simulation is difficult because each unit has a different topological from

- Display, and plotting can be expensive, particularly for high quality, colour and crosshatching
- The technology is expensive, particularly for the more sophisticated software and hardware
- Spatial analysis and filtering within polygons are impossible

**Raster methods**

**Advantages :-**

- Simple data structures
- The overlay and combination of mapped data with remotely sensed data is easy
- Various kinds of spatial analysis are easy
- Simulation is easy because each spatial unit has the same size and shape
- The technology is cheap and is being energetically develop

**Disadvantages :-**

- Volumes of graphic data
- The use of large cells to reduce data volumes means that phenomenologically Recognizable structures can be lost and there can be a serious loss of information
- Crude raster maps are considerably less beautiful than maps drawn with fine lines
- Network linkages are difficult to establish
- Projection transformation are time consuming unless spatial algorithms or hardware are used.

The problem of raster or vector disappears once it is realized that both are valid methods for representing spatial data, and that both structure are interconvertible. Conversion from vector to raster is the simplest and there are many well know algorithms. Vector to raster conversions are now performed automatically in many display screens by inbuilt microprocessors. The reverse operation raster to vector, is also well understood (Pavlidis lists four algorithms for thinning bands of pixel to lines), but it is a much more complex operation that is complicated by the need to reduce the number of co-ordinates in the resulting lines by a process know as weeding.

**Suggestions For The Use Of Raster And Vector Methods**

1. Use VECTOR data structure for data archiving phenomenologically structured data (e.g. soil areas, land use units, etc.).
2 Use VECTOR methods for network analyses, such as for telephone networks, or transport network analysis.
3. Use VECTOR data structure and VECTOR display methods for the highest quality line drawing.
4. Use RASTER methods for quick and shear map over lay, map combination and spatial analysis.

5. Use RASTER methods for simulation and modeling when it is necessary to work with surfaces.
17
6. Use RASTER and VECTOR in combination for plotting high quality lines in combination with efficient area filling in colour. the lines can be held in VECTOR format and the raster filling in compact RASTER structures such as run length codes or quadtrees.
7. Preferably use compact VECTOR data structure for digital terrain models, but don't neglect altitude matrices.
8. Use RASTER -VECTOR and VECTOR - RASTER algorithms to convert data to the most suitable form for a given analysis or manipulation.
9. Remember that DISPLAY systems can operate either in RASTER or VECTOR modes independent of the DATA STRUCTURES that are used to store and manipulate the data.