TRAINING COURSE
ON

HYDROLOGICAL MODELING AND GIS

(MAY 26 TO JUNE 06, 2014)

FOR

UNFAO & Ministry of Energy and Water, Afghanistan

LECTURE NOTE
ON

INTRODUCTION TO GIS

By

R. D. GARG
IIT ROORKEE

NATIONAL INSTITUTE OF HYDROLOGY
JALVIGYAN BHAWAN
ROORKEE – 247 667 (UTTARAKHAND)

INTRODUCTION TO GEOGRAPHICAL INFORMATION SYSTEMS

INTRODUCTION

Experts of many disciplines can be classified as users of spatial data. Next to geodesists and cartographers some of these experts are geographers, geologists, soil scientists, environmentalists, planners and remote sensing scientists. When spatial data are used in an information system one tends to speak of a spatial information system. When such a system has grown fully, it integrates the disciplines as represented.

In the period when the computer became an important tool in each separate discipline, the main purpose of its use was data inventory. Spatial data were observed, collected, classified and stored. In the next phase, the experts who collected the data became interested in spatial analysis of those data. Today the need of spatial analysis crosses the borders of the separate disciplines, and people want an over all approach. The use of spatial information systems offer the opportunity to fulfill these needs.

GIS as in Integrating Technology

Geographic information systems have served an important role as an integrating technology. Rather than being completely new, GIS have evolved by linking a number of discrete technologies into a whole that is greater than the sum of its parts. GIS have emerged as very powerful technologies because they allow geographers to integrate their data and methods in ways that support traditional forms of geographical analysis, such as map overlay analysis as well as new types of analysis and modelling that are beyond the capability of manual methods. With GIS it is possible to map, model, query, and analyze large quantities of data all held together within a single database.

The important of GIS as in integrating technology is also evident in its pedigree. The development of GIS has relied on innovations made in many different disciplines : Geography, Cartography, Photogrammetry, Remote Sensing, Surveying, Geodesy, Civil engineering, Statistics, Computer Science, Operations Research, Artificial Intelligence Demography, and many other branches of the social science, natural sciences, and engineering have all contributed. Indeed, some of the most interest applications of GIS technology discussed below draw upon this interdisciplinary character and heritage.

TERMINOLOGY

The term spatial information system can be seen as the general discipline for those information systems which deals with spatial data. Geographical Information System (GIS) is commonly used in the context of geo-base information system, natural resources information system, geo-data system, geographical data system. They all stress the spatial analysis capacities of their system.

Besides GIS there is also LIS (Land Information System). A LIS is oriented on storing, retrieval, updating and presenting land data. The main geographic elements of a LIS are building, land parcels, topographic elements, cables and pipelines. The information connected to these elements are mainly technical data, juridical data, value, use and relations with other information systems. A LIS is more detailed than a GIS. In a LIS the information of a land parcel and information about the owner and the juridical relation between them can be obtained. Sometimes a LIS is considered as a special implementation of a GIS. Thus a LIS and a GIS have both special characteristics, and together they cover the field of spatial information systems.

The definition given by various authors are as follows :

- A spatial data handling system" (Marble et al, 1983).
- "A computer - assisted system for the capture, storage, retrieval, analysis and display of spatial data, within a particular organization" (Clarke, 1986).
- "A powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world" (Burrough, 1987).
- "An internally referenced, automated, spatial information system" (Berry, 1986).
- "A system, which uses a spatial data base to provide answers to queries of a geographical nature" (Goodchild, 1985).

Geographic Information Systems : A Generic Definition

GIS is a special-purpose digital database in which a common spatial coordinate system is the primary means of reference.

Comprehensive GIS require a means of:

1. Data input, from maps, aerial photos, satellites, surveys, and other sources.
2. Data storage, retrieval, and query
3. Data transformation, analysis and modeling, including spatial statistics
4. Data reporting, such as maps, reports and plans

The Appeal and Potential of GIS:

The great appeal of GIS stems from its ability to integrate great quantities of information about the environment and to provide a powerful repertoire of analytical tools to explore this data. The example above displayed only a few map layers pertaining to urban transportation planning. The layers included would be very different if the application involved modeling the habitat of an endangered species or the environmental consequences of leakage from a hazardous materials site.

Imagine the potential of a system in which dozens or hundreds of maps layers are arrayed to displayed information about transportation networks, hydrographic, population characteristics, economic activity, potential jurisdictions, and other characteristics of the natural and social environments. Such a system would be valuable in a wide range of situations for urban planning, environmental resource management, hazards management, emergency planning, or transportation forecasting, and so on. The ability to separate information in layers, and then combine it with other layers of information is the reason why GIS hold such great potential as research and decision making tools.

Application Areas

GIS are now used extensively in government, business, and research for a wide range of applications including environmental resource analysis, land use planning, locational analysis, tax appraisal, utility and infrastructure planning, real estate analysis, marketing and demographic analysis, habitat studies, and archaeological analysis.

One of the first major areas of application was in natural resources management, including management of

Wildlife habitat
Wild and scenic rivers
Recreation resources
Floodplains
Wetlands,
Agricultural lands
Aquifers
Forests

One of the largest areas of application has been in facilities management. Uses for GIS in this area have included.

Locating underground pipes and cables,
Balancing loads in electrical networks,
Planning facility maintenance,
Tracking energy use

Local, state, and federal governments have found GIS particularly useful in land management. GIS has been commonly applied in areas like

Zoning and subdivision planning
Land acquisition
Environmental impact policy
Water quality management
Maintenance of ownership

More recent and innovative uses of GIS have used information based on street networks. GIS has been found to be

Particularly useful in

Address matching
Location analysis or site selection,
Development of evacuation plans.

The range of applications for GIS is growing, as system become more efficient, more common, and less expensive.

The components of a geographical information system

Geographical information systems have three important components : computer hardware, set of application software modules, and a proper organizational context. These three components need to be in balance if the system is to function satisfactorily.

1. Computer hardware

The general hardware components of a geographical information system are presented in Fig.2. The computer or central processing unit (CPU) is linked to a disk drive storage unit, which provides space for storing data and programs. A digitizer or other device is used to convert data from maps and documents into digital form and send them to the computer. A plotter or other kind of display device is used to present the results of the data processing, and a tape drive is used for storing data or programs on magnetic tape, or for communicating with other systems. Inter-computer communication can also take place via a networking system over special data lines, or over telephone lines using a device known as a `modem'. The user controls the computer and the peripherals (a general term for plotters, printers, digitizers, and other apparatus linked to the computer) via a visual display unit (VDU), otherwise known as a terminal. The user's terminal might itself be a microcomputer, or it might incorporate special hardware to allow maps to be displayed quickly.

2. GIS software modules

The software package for a geographical information system consists of five basic technical modules (Fig.3). These basic modules are sub-systems for :

a) Data input and verification;
b) Data storage and database management;
c) Data output and presentation;
d) Data transformation
e) Interaction with the user

a) Data input : Covers all aspects of transforming data captured in the form of existing maps, field observations, and sensors (including aerial photography, satellites, and recording instruments) into a compatible digital form. A wide range of computer tools is available for this purpose, including the interactive terminal or visual display device (VDU), the digitizer, lists of data in text files, scanners (in satellites or aeroplanes for direct recording of data or for converting maps and photographic images) and the devices necessary for recording data already written on magnetic media such as tapes, drums, and disks.

b) Data storage and database management : Concerns the way in which the data about the position, linkages (topology), and attributes of geographical elements (points, lines, and areas representing objects on the earth's surface) are structured and organized, both with respect to the way they must be handled in the computer and how they are perceived by the users of the system. The computer program used to organize the database is known as a Database Management System (DBMS).

c) Data output and presentation: Embraces two classes of operation, namely (a) transformations needed to remove errors from the data sets, and (b) the large array of analysis methods that can be applied to the data in order to achieve answers to the questions asked of the GIS. Transformations can operate on the spatial and the non-spatial aspects of the data, either separately or in combination. Many of these transformations, such as those associated with scale-changing, fitting data to new projections, logical retrieval of data, and calculation of areas and perimeters, are of such a general nature that one should expect to find them in every kind of GIS in one form or another. Other kinds of manipulation may be extremely application-specific, and their incorporation into a particular GIS may be only to satisfy the particular users of that system.

d) The last module in the list for geographical information systems mentioned above, that of interaction with the user - query input - is absolutely essential for the acceptance and use of any information system. Certainly it is an aspect that until relatively recently has received less attention than it deserves. It is only in the last few years that the average user has been able to make direct contact with the computer other than via the impersonal and unforgiving media of punched paper tapes and cards handed in to the computing centre. The widespread introduction of the personal computer and of programs that are operated by commands chosen from a menu (a list), or that are initiated by a response to requests in an English-like command language of verbs, nouns, and modifiers has broken down the barriers that once frightened many a would-be computer user away for life.

3. The organizational aspect of GIS

The five technical sub-systems of GIS govern the way in which geographical information can be processed but they do not of themselves guarantee that any particular GIS will be used effectively. In order to be used effectively, the GIS needs to be placed in an appropriate organizational context. It is simply not sufficient for an organization to purchase a computer and some software and to hire or retrain one or two enthusiastic individuals and then to expect instant success. Just as in all organizations dealing with complex products, as in manufacturing industry, new tools can only be used effectively if they are properly integrated into the whole work process and not tackled on as an afterthought. To do this properly requires not only the necessary investment in hardware and software, but also in the retraining of personnel and managers to use the new technology in the proper organizational context.

Geographical data in the 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 DATA

Points, lines, and areas

All geographical data can be reduced to three basic topological concepts - the point, the line, and the area. Every geographical phenomenon can be represented by a point, line or area plus a label saying what it is. So an oil well could be represented by a point entity consisting of a single XY coordinate pair and the label `oil well'; a section of railway line could be represented by a line entity consisting of a starting XY coordinate and an end XY coordinate and the label `railway'; a floodplain could be represented by an area entity covering a set of XY coordinates plus the label `floodplain'. The labels could be the actual names as given here, or they could be numbers that cross-reference with a legend, or they could be special symbols

The essential features of any data storage system are that they should be able to allow data to be accessed and cross-referenced quickly. There are several ways of achieving this, some of which are more efficient than others. Unfortunately, there seems to be no one `best' method that can be used for all situations. This explains in part the massive investment in labour and money in effective database management systems, which are the computer programs that control data input, output, storage, and retrieval from a digital database.

DATA BASE

Data are raw material from which every land information system is built. They are gathered and assembled into records and files.

A data base is a collection of data that can be shared by different users. It is a group of records and files that are organized so that there is little or no redundancy.

Database structures

A database consists of data in many files. In order to be able to access data from one or more files easily, it is necessary to have some kind of structure or organization. Three main kinds of database structure are commonly recognised, termed Hierarchical, network, and relational.

Flat Files and Spreadsheets

A flat or spreadhseet is a simple method for storing data. All records in this database have the same number of "fields". Individual records have different data in each field with one field serving as a key to locate a particular record. For example, your social security number may be the key field in a record of your name, address, phone number, sex, ethnicity, place of birth, date of birth, and so on. For a person, or a tract of land there could be hundreds of fields associated with the record. When the number of fields becomes lengthy a flat file is cumbersome to search. Also the key field is usually determined by the programmer and searching by other determinants may be difficult for the user. Although this type of database is simple in its structure, expanding the number of fields usually entails reprogramming. Additionally, adding new records is time consuming, particularly when there are numerous fields. Other methods offer more flexibility and responsiveness in GIS.

Flat File Structure

Tax No.	**Street Address**	**Sub div.**	**Block/ Lot**	**Owner 1**	**Owner 2**	**Improved**	**GC Warbler**
234	10 Lone Oak	RobRoy	A/116	Verdi, G.	Rossini, G.	Yes	No
235	12 Lone Oak	RobRoy	A/118	Wagner, R.	Weber, C.	No.	Yes
236	101 Madrone	Live Oak	B/14	Hendrix, J.	Morrison, J.	Yes	Yes

All individual property data is in one file. The tax No. is required to search the database.

Hierarchical data structure

When the data have a parent/child or one-to-many relation, such as soil series within a soil family, or pixels within a region, hierarchical methods provide quick and convenient means of data access.

Hierarchical systems assume that each part of the hierarchy can be reached using a key (a set of discriminating criteria) that fully describes the data structure.

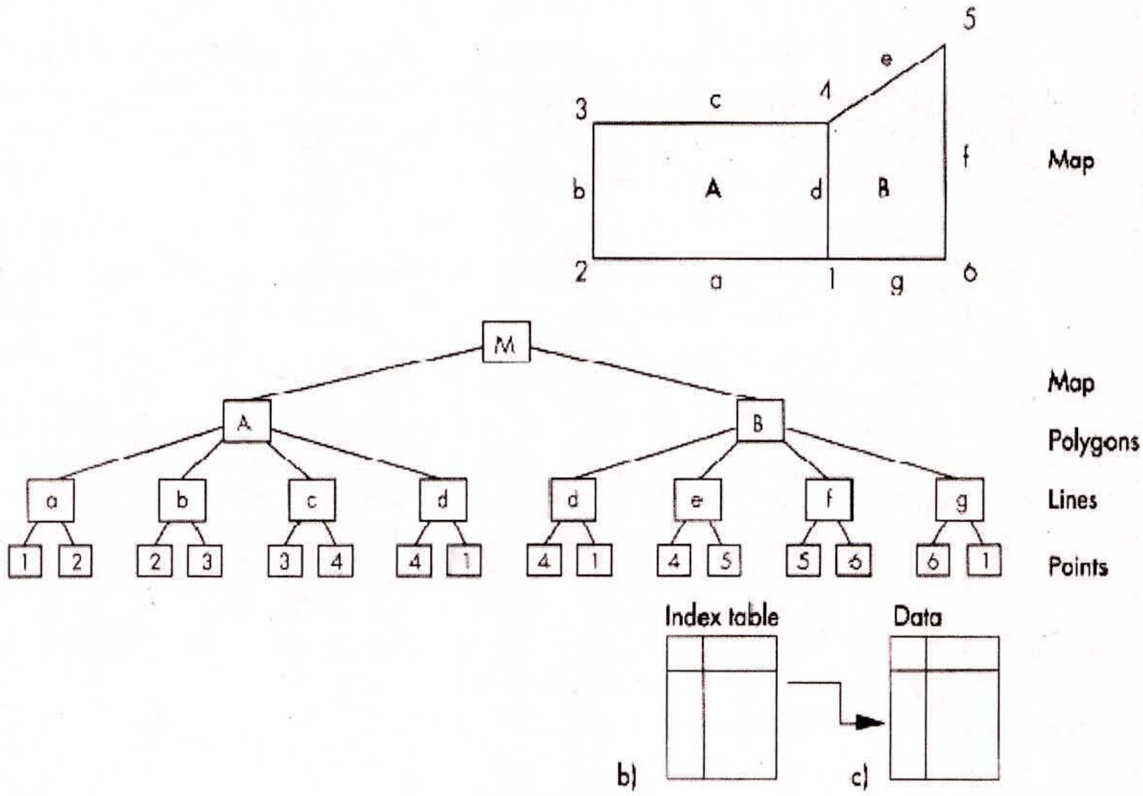


Figure: Map data stored in an hierarchical database system. Note the double storage of lines and points

Hierarchical File Structure

Tax No.	Street Address
234	10 Lone Oak
235	12 Lone Oak
236	101 Madrone

Tax No	Subdiv.	Block/Lot
234	RobRoy	A/116
235	RobRoy	A/118
236	LiveOak	B/14

Tax No	Improved	GC Warbler
234	Yes	No
235	Yes	Yes
236	No	Yes

Tax No. is the "pointer" or link to different records.

Tax No	Owner	Address
234	Verdi, G.	10 Traviata, Macbeth, TX
234	Rossini, G.	Box 7, Siviglia, TX
235	Wagner, R.	12 Lone Oak
235	Weber, C.	12 Lone Oak
236	Hendrix, J.	201 Haze Rd, Purple, TX
236	Morrison, J.	33 Light Fire Tr, End, TX

Hierarchical systems have the advantage that they are easy to understand and they are easy to update and expand. Data access via the keys is easy for key attributes, but unfortunately is very difficult for associated attributes. Consequently, hierarchical systems are good for data retrieval if the structure of all possible queries can be known before hand. This is commonly the case with bibliographic, bank or airline retrieval systems.

Further disadvantages of hierarchical database structures are that large index files have to be maintained, and certain attribute values may have to be repeated many times, leading to data redundancy, which increases storage, and access costs.

Network systems

In hierarchical systems, travel within the database is restricted to the paths up and down the taxonomic pathways. In many situations much more rapid linkage is required, particularly in data structures for graphics features where adjacent items in a map or figure need to be linked together even though the actual data about their coordinates may be written in very difficult parts of the database. Network systems fulfill this need.

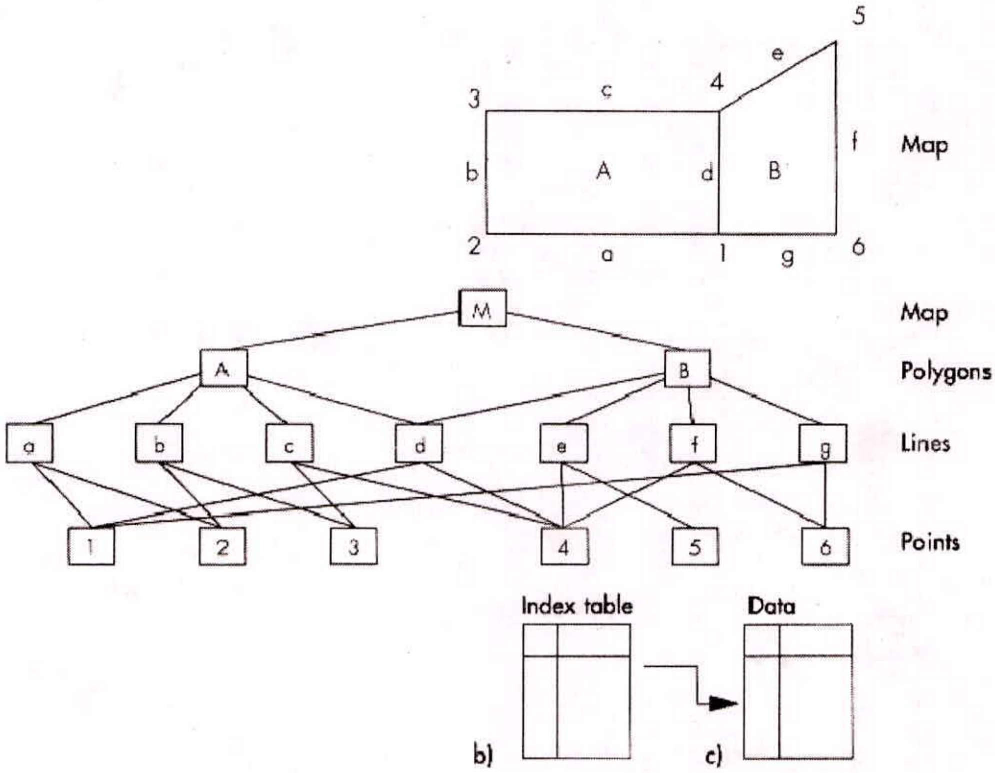


Figure. The storage structure in a network-based database system. a) shows a map section; b) shows polygons stored in the network; c) is a pointer table which specifies where the map data (d) are in fact stored

Relational database structures

The relational database structure in its simplest form stores no pointers and has no hierarchy. Instead, the data are stored in simple records, known as tuples, containing an ordered set of attribute values that are grouped together in two-dimensional tables, known as relations. Each table or relation is usually a separate file. The pointer structures in network models and the keys in hierarchical structures are replaced by data redundancy in the form of identification codes that are used as unique keys to identify the records in each file.

Data are extracted from a relational database through a procedure in which the user defines the relation that is appropriate for the query. This relation is not necessarily already present in the existing files, so the controlling program uses the methods of relational algebra to construct the new tables.

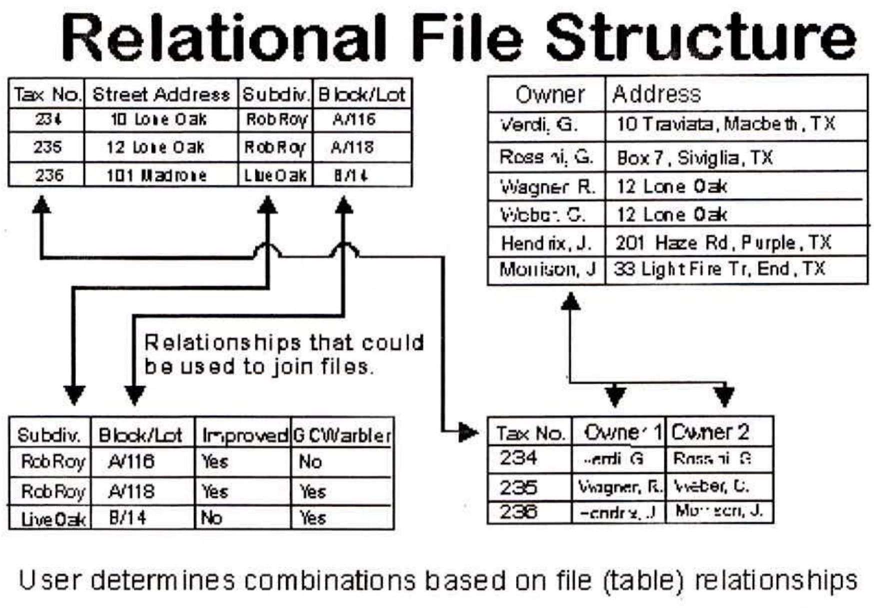


Figure: A relations database. Each field in a table can be the key to locating data in another table

There are no pointers in the data tables. Internal index tables are used to direct inter-table communications, which means that the system must open at least one index table for each connection between data tables. This can result in relatively large databases and slow access. Moreover, as there are no pointers, searches for records or fields must be sequential through the tables. The result is that relational database systems are inherently slower than hierarchical or network database systems.

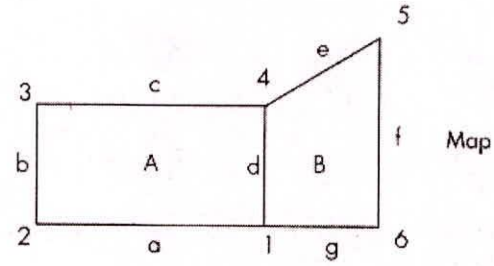
Relational database have the great advantage that their structure is very flexible and can meet the demands of all queries that can be formulated using the rules of Boolean logic and of mathematical operations. They allow different kinds of data to be searched, combined, and compared. Addition or removal of data is easy too, because this just involves adding or removing a tuple. The disadvantage of relational databases is that many of the operations involve sequential searches through the files to find the right data to satisfy the specified relations. This can involve a considerable amount of time with large databases, even on fast computers.

Representation of Topological Data-

There are at least two fundamental ways of representing topological data which can be summarized as follows:

Raster representation

Set of cells located by coordinates, each cell is independently addressed with the value of an attribute.



Map

A	B

Polygons

A	a	b	c	d
B	c	e	f	g

Lines

a	1	2
b	2	3
c	3	4
d	4	1
e	4	5
f	5	6
g	6	1
d	4	3

Points

1	X_1	Y_1
2	X_2	Y_2
3	X_3	Y_3
4	X_4	Y_4
5	X_5	Y_5
6	X_6	Y_6

Figure: Examples of map data stored in a relations database

Vector representation

Three main geographical entities, points, lines and areas; points are similar to cells, except they do not cover areas; lines and areas are sets of interconnected coordinates that can be linked to given attributes.

Note that there is no necessary or unique connection between the raster or vector structure of the geographical database and the structure of the devices used to display the raster and vector data, although this is very often the case. For example, most modern interactive computer-aided design and mapping systems, work with vector-structured databases but use colour raster displays and vector plotters.

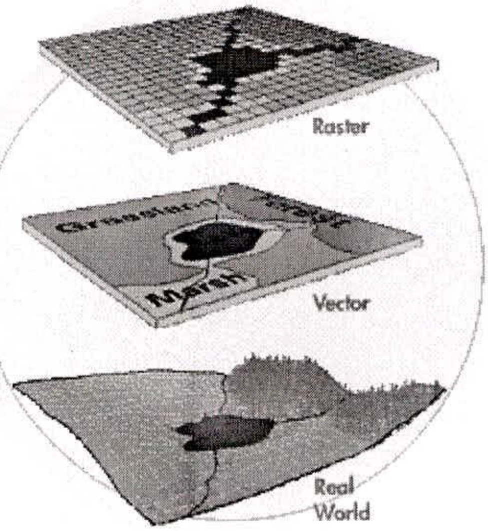


Figure : Vector and Raster Representations

RASTER DATA STRUCTURES :

The Raster View of the World :

A raster based system displays, locates, and stores graphical data by using a matrix or grid of cells. A unique reference coordinate represents each pixel either at a corner or the centroid. In turn each cell or pixel has discrete attribute data assigned to it. Raster data resolution is dependent on the pixel or gird size and may vary from sub-meter to many kilometers. Because these data are two-dimensional, GISs store various informations such as forest cover, soil type, land use, wetland habitat, or other data in different layers. Layers are functionally related map features. Generally, raster data requires less processing than Vactor data, but it consumes more computer storage data space. Scanning remote sensors on satellites stores data in raster format. Digital terrain models (DTM) and digital elevation models (DEM) are examples of raster data.

Map Overlays :

Since each cell in a two dimensional array can only hold one number, different geographical attributes must be represented by separate sets of Cartesian arrays, known as 'overlays'. In its simplest form, the overlay concept is realized in raster data structures by stacking two dimensional arrays. This results in the three dimentional structure.

Compact Methods for Storing Raster Data :

When each cell has a unique value it takes a total of n rows X m columns X 3 values (X, Y coordinates and attribute value) to encode each overlay. If sets of cells within a polygon or a mapping unit all have the same value, however, it is possible to effect considerable savings in the data storage requirements for the raster data, providing of course that the data structures are properly designed. There are four main ways in which compact storage can be achieved.

(i) Chain Codes

Chain codes provide a very compact way of storing a region representation and they allow certain operations such as estimation of areas and perimeters, or detection of sharp turns and concavities to be carried out easily. On the other hand, overlay operations such as union and intersection are difficult to perform without returning to a full grid representation. Another disadvantage is the redundancy introduced because all boundaries between the regions will be stored twice.

(ii) Run – length codes :

Run-length codes allow the points in each mapping unit to be stored per row in terms, from left to right, of a begin cell and an end cell. These are useful in reducing the volume of data that need to be input to a simple raster database.

(iii) Block codes :

The idea of run-length codes can be extended to two dimensions by using square blocks to tile the area to be mapped. The data structure consists of just three numbers, the origin (the center or bottom left) and radius of each square. Both run-length and block codes are clearly most efficient for large simple shapes and so for small complicated areas that are only a few times larger than the basic cell. For some operations, data stores in block run-length codes must be converted to simple raster format.

(iv) Quadtree :

The fourth method of more compact representation is based on successive dividion of the 2" x 2" array into quadrants. A region is tiled by subdividing the array step by step into quadrants and nothing which quadrants are wholly contained with the region. The lowest limit of division is the single pixel. Show the successive division of the one region of fig. 10 into quadrant blocks. This block structure can be described by a tree of degree 4, know as the quadtree, the quadtree representation of is given in fig quadtrees are resource information systems.

If each cell represents a potentially different value, then the simple N x N structure is difficult to improve upon. Its limitations are largely related to the volume of data and size of memory required. When 'regions' (i.e. area of uniform value) are present, as is assumed to be considerably reduced by using length codes appear to be most efficient when the pixel size is large with respect to the areas of the region being displayed and sorted, as resolution improves and pixel numbers per region increase however, block codes and quadtrees become increasingly attractive. The quadtree representation has the added advantage of variable resolution. The ease of subsequent processing various with the data structure used.

VECTOR DATA STUCTURES :

The vector representation of an object is an attempt to represent the object as exactly as possible. The coordinate space is assumed to be continuous, not quantised as with the raster space, allowing all positions, lengths, and dimensions to be defined precisely, a range of vector structures used in geographical information systems 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 coordinate 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.

Line Entities :

Line entities can be defined as all liner features built up of straight-line segments made up of two or more coordinates. The simplest line required the storage of a being point and an end point (two XY coordinate pairs) plus a possible records indicating the display symbol to be used.

An 'arc' a 'chain' or a 'string' is a set of n XY coordinates pairs describing a continuous complex line, the shorter the line segments, and the larger the number of XY coordinates 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 coordinate when the line data are sent to the display device.

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

Area Entities :

Areas of polygons (or regions) can be represented in various ways in a vector database. The simplest way to represent a polygon is an extension of the simple chain, i.e.

to represent each polygon as a set of XY coordinates on the boundary. The names or symbols used to tell the user what each polygon is are then held as a set of simple text entities. While this method has the held as a set simple text entities. While this method has the advantages of simplicity it has many disadvantages : (a) Lines between adjacent ploygons must be digitized and stored twice, (b) there is no neighborhood information, (c) Islands are impossible except as purely graphical constructions, and (d) Three are no easy ways to check if the topology of the boundary is correct or whether it is incomplete (dead ends) or makes toplogically inadmissible loops ('weird polygons').

The simple polygon structure can be extended such that each polygon is represented by a number of chains.

The choice between Raster and Vector

The raster and vector methods for spatial data structures are distinctly different approaches to modelling 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 required huge computer memories to store and process image at the leel 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.

The problem of raster or vector disappears once it is realized that both are valid methods for representing spatial data, and that both structures are interconvertible. Conversion from vector to raster is the simplest and there are many well-known algorithms (e.g. Pavelidis 1982). 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 pixels to lines), but it is a much more complex operation that is complicated by the need to reduce the numbers of coordinates in the resulting lines by a process known as weeding.

Comparison of vector and raster methods

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 form
- Display and plotting can be expensive, particularly for high quality, colour and cross-hatching
- 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 developed

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 special algorithms or hardware are used.

DATA INPUT and OUTPUT

For a GIS to be useful it must be capable of receiving and producing information in an effective manner. The data input and output functions are the means by which a GIS communicates with the world outside. The objective in defining GIS input and output requirements is to identify the mix of equipment and methods needed to meet the required level of performance and quality. No one device or approach is optimum for all situations.

DATA INPUT

It is the procedure of encoding data into a computer-readable form and writing the data to the GIS database. Data entry is usually the major bottleneck in implementing a

GIS. The initial cost of building the database is commonly 5 to 10 times to cost of the GIS hardware and software. The creation of an *accurate* and *well-documented* database is critical to the operation of the GIS. Accurate information can only be generated if the data on which it is based were accurate to begin with. Data quality information includes the date of collection, the positional accuracy, completeness, and the method used to collect and encode the data.

There are two types of data to be entered into a GIS: **Spatial data and the associated non-spatial attribute data**.

The spatial data represents the geographical location of the features. The non-spatial attribute data provide descriptive information like the name of a street, salinity of the lake or the type of tree stand. The non-spatial attribute data must be logically attached to the features they describe.

The methods of data acquisition can be classified as (*Bill and Fritsch*):

- Primary methods (acquisition of data from the object itself)
- Secondary methods (data acquisition from existing analogue or digital sources).

DATA OUTPUT

Output is the procedure by which information from the GIS is presented in a form suitable to the user. Data are output in one of three formats: Hardcopy, Softcopy and electronic. Hardcopy outputs are permanent means of display. The information is printed on paper, mylar, photographic film or other similar materials. Softcopy output is in the format viewed on a computer monitor. Softcopy outputs are used to allow operator interaction and to preview data before final output. A Softcopy output can be changed interactively but the view is restricted by the size of the monitor.
The hardcopy output takes longer to produce and requires more expensive equipment. However, it is a permanent record. Output in electronic formats consists of computer-compatible files. The output from GIS does not have to be a map in fact, many GIS are designed with poor map output capabilities.