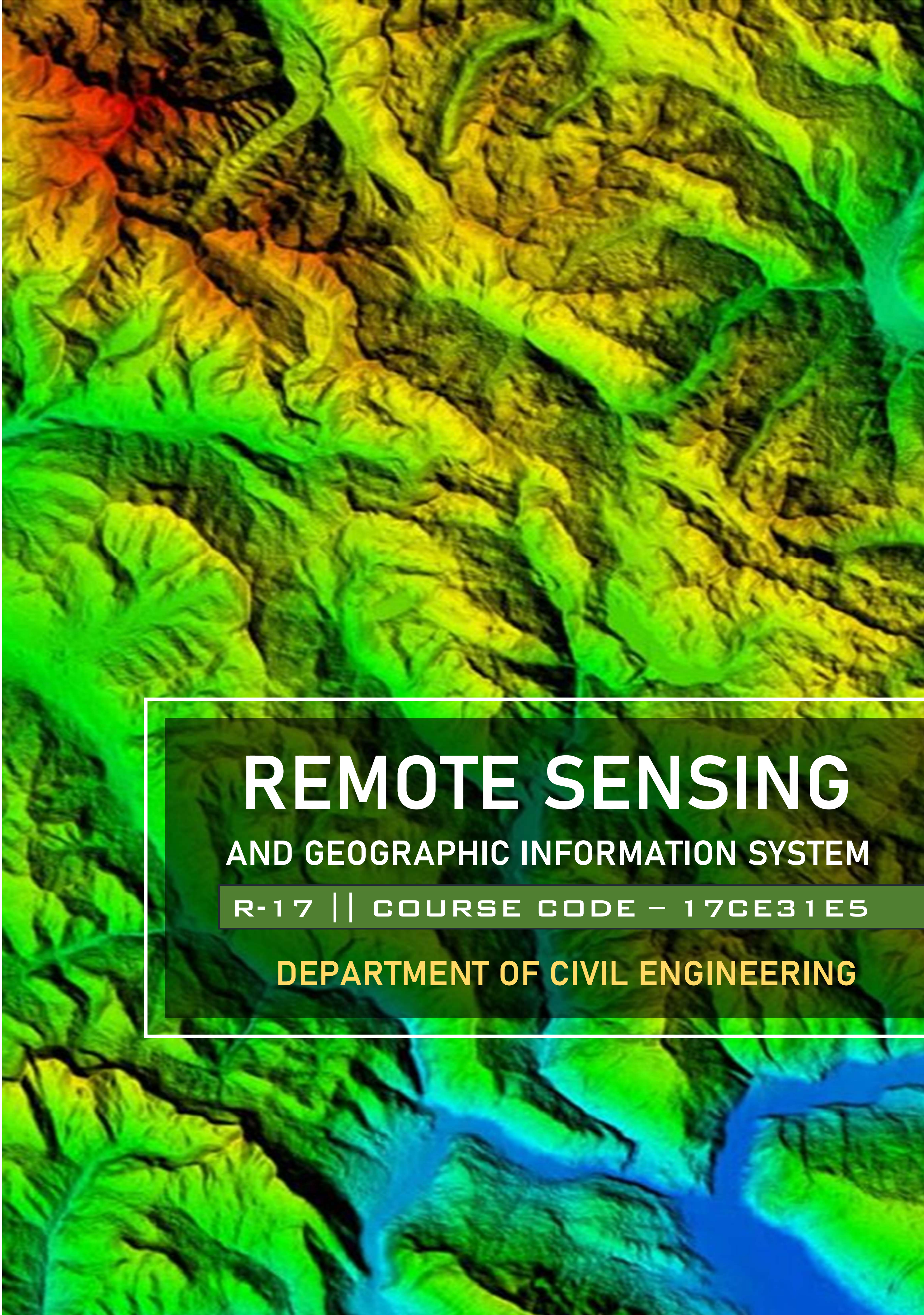


REMOTE SENSING

AND GEOGRAPHIC INFORMATION SYSTEM

R-17 || COURSE CODE – 17CE31E5

DEPARTMENT OF CIVIL ENGINEERING

MODULE – V

GEOGRAPHICAL INFORMATION SYSTEM

Course objective	To understand the elements of GIS.
Content	**GEOGRAPHICAL INFORMATION SYSTEM:** Basic Principles – Definition – Components – Data structures – Raster and vector formats – Functioning of GIS – Data Input – Data manipulation – Data retrieval – Data analysis – Data display – Data base management systems.
Course outcome	Understand the basic concepts of Geographical Information System

Introduction

- GIS stands for Geographical Information System. The geographical information system (GIS) consists of two distinct disciplines, namely geography and information system.
- Geography is the scientific study of geo-spatial pattern and process. It seeks to identify and account for the location and distribution of human and physical phenomena on the earth's surface.
- Emphasis in geography is placed upon the organization and arrangement of phenomena, and upon the extent to which they vary from place to place and time to time. It is the characteristics of space as a dimension, rather than the properties of phenomena which are located in space, that is of central and overriding concern.
- Geography aims to develop general rather than unique explanations. It proceeds from the assumption that there is basic regularity and uniformity in the location and occurrence of phenomena and that this order can be identified and accounted by geographical analysis. In examining spatial structure, geography focuses upon those distributional characteristics that are common wide range of phenomena.
- Information system most often refers to a system containing electronic records, which involves input of source documents, recording on electronic media, and output of records, along with related documentation and any indexes.
- The information system can be defined as an interactive combination of people, computer hardware and software, communications devices, and procedures

designed to provide a continuous flow of information to the people who need information to make decisions or perform analysis.

- The GIS is a computer-based information system used to digitally represent and analyze the geospatial data of geographic data. Geospatial means the distribution of something in a geographic sense; it refers to entities that can be located by geographic coordinate system.

- 'Every object present on the earth can be georeferenced', is the fundamental key of associating any database to GIS. The GIS is a particular form of information system applied to geographical data. A system is a group of connected entities and activities which interact for a common purpose.
- Earlier we had paper maps, which were very colorful, but not modifiable. Then, the computer revolution took place, where the maps were digitalized and stored in digital format.

DEFINITION OF GIS

- The GIS is an information system designed to work with data referenced by spatial/geographical coordinates.
- **GIS is defined as** an information system that is used to input, store, retrieve, manipulate, analyze and output geographically referenced data or geospatial data, in order to support decision making for planning and management of land use, natural resources, environment, transportation, urban facilities, and other administrative records.
- A geographic information system is a facility for preparing, presenting, and interpreting facts that pertain to the surface of the earth.
- A computer based system that provides four sets of capabilities to handle georeferenced data: data input, data management, manipulation and analysis, and data output.

OBJECTIVES OF GIS:

Some of the objectives of GIS are as follows:

- Maximizing the efficiency of planning and decision making.
- Integrating information from multiple sources.
- Facilitating complex querying and analysis.
- Eliminating redundant data and minimizing duplication.

COMPONENTS OF GIS:

- The GIS constitutes of five key components, namely, hardware, software, procedure, data, and users as shown in figure.

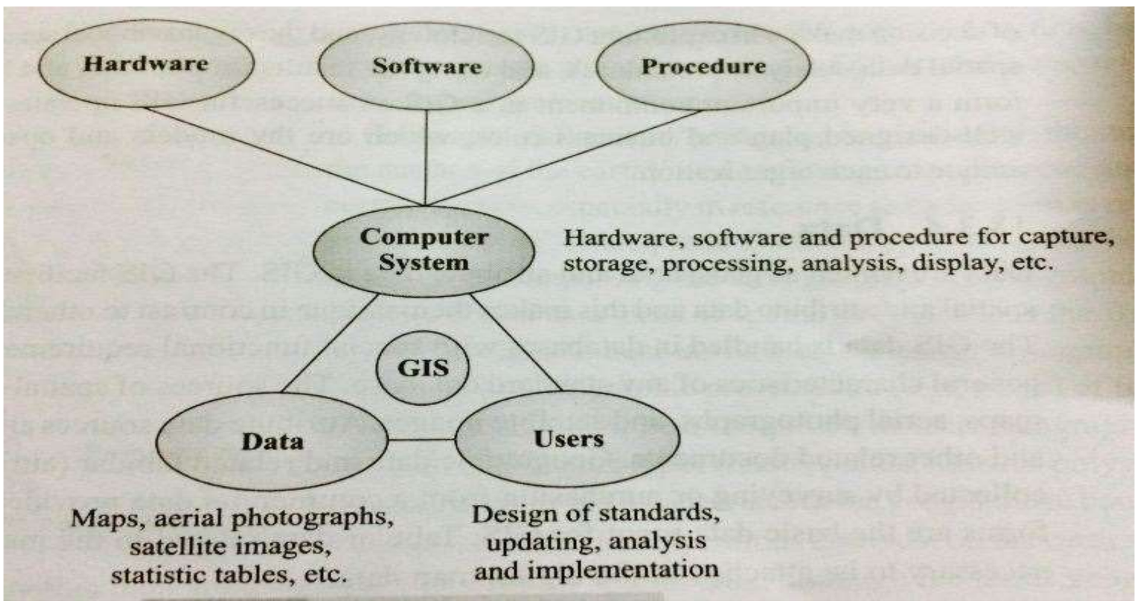


Fig-1. Components of GIS

HARDWARE:

It consists of computer hardware system on which the GIS software runs. The GIS run on the whole spectrum of computer systems ranging from portable personal computers to multi user supercomputers. The hardware of GIS consists of input devices such as digitizers, scanners and GPS receivers, the storage devices such as magnetic tapes and disks, CD ROMs and other optical disks, central processing units, and the output devices such as display devices, printers and plotters.

SOFTWARE:

Software is at the heart of a GIS system. The GIS software must have the basic capabilities of data input, storage, transformation, analysis and providing desired outputs. The interfaces could be different for different software's. The GIS software's

being used today belong to either of the category –proprietary or open source. ArcGIS by ESRI is the widely used proprietary GIS software. Others in the same category are MapInfo, Micro station, Geomedia etc. The development of open source GIS has provided us with freely available desktop GIS such as Quantum, uDIG, GRASS, Map Window GIS etc., GIS software's.

PROCEDURE:

A computer system for GIS consists of hardware, software and procedures designed to support the data capture, storage, processing, analysis, modeling, and display of geospatial data. Besides the technical components like hardware, software and databases, institutional framework and policies are also important for a functional GIS. A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

DATA:

Data are named as geospatial and attribute data in GIS. The GIS facilitates integration of spatial and attribute data and this makes them unique in contrast to other database systems. The GIS data is handled in databases with special functional requirements as well as the general characteristics of any standard database. The sources of spatial data are digitalized maps, aerial photographs and satellite images. Attribute data sources are statistical tables and other related documents. The digital map forms are the basic data input for GIS.

USERS:

The role of the user are to select pertinent information to set necessary standards, to design cost-efficient updating schemes, to analyze GIS outputs for relevant purposes and plan the implementation. Most definition of GIS focuses on the hardware, software, data and analysis components. The GIS projects range from small research applications, where one user is responsible for design and implementation and output,

to international corporate distributes systems, where different type of users interact with the GIS in many different levels and ways.

SPATIAL DATA:

Spatial data also known as geospatial **data** or geographic information it is the **data** or information that identifies the geographic location of features and boundaries on Earth, such as natural or constructed features, oceans, and more. **Spatial data** is usually stored as coordinates and topology, and is **data** that can be mapped.

ATTRIBUTE DATA:

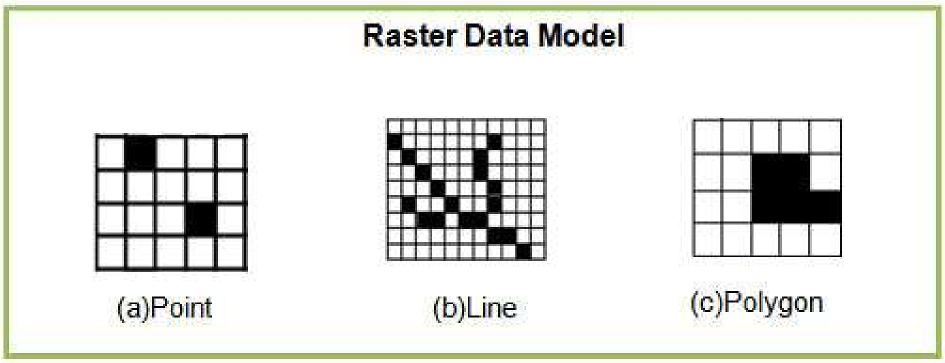
Attribute data is qualitative **data** that can be counted for recording and analysis. **Attribute data** is not acceptable for production part submissions unless variable **data** cannot be obtained.

DATA STRUCURES

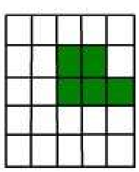
- Spatial **Data Structures**. Spatial **data** types provide the information that a computer requires to reconstruct the spatial **data** in digital form.
- A data model describes in an abstract way how data is represented in an information system or in a database management system. The manner in which data is generally organized in a database management system is sometimes also called a database model.
- There are two fundamental approaches to the representation of the spatial component of geographic information: vector model and raster model.

RASTER DATA STRUCTURE:

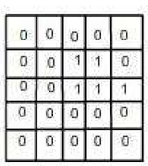
- The raster data model is commonly associated with the field conceptual model. Here, geographic space is represented by array of cells or pixels which are arranged in rows and columns. Each pixel has a value that represents information. The value can be in the form of integer, floating points or alphanumeric.
- A point can be represented by a single pixel in raster model. A line is a chain of spatially connected cells with the same value. Similarly, a water body in raster data is represented as a set of contiguous pixels having same value that represents a homogeneous area.



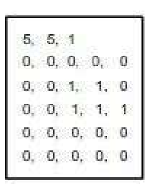
- In a simple raster data structure the geographical entities are stored in a matrix of rectangular cells. A code is given to each cell which informs users which entity is present in which cell. The simplest way of encoding a raster data into computers can be understood as follows:



(a) Entity model: It represents the whole raster data. Let us assume that the raster data belongs to an area where land is surrounded by water. Here a particular entity (land) is shown in green color and the area where land is not present is shown by white.



(b) Pixel values: The pixel value for the full image is shown. Cells having a part of the land are encoded as 1 and others where land is not present are encoded as 0.



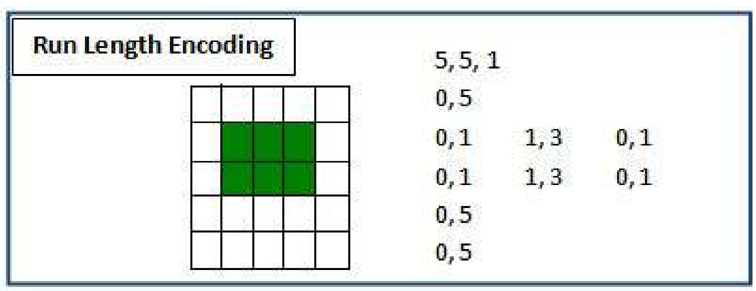
(c) File structure: It demonstrates the method of coding raster data. The first row of the file structure data tells that there are 5 rows and 5 columns in the image, and 1 is the maximum pixel value. The subsequent rows have cells with value as either 0 or 1 (similar to pixel values).

- The huge size of the data is a major problem with raster data. An image consisting of twenty different land-use classes takes the same storage space as a similar raster map showing the location of a single forest. To address this problem many data compaction methods have been developed as discussed below:

Run Length Encoding(RLE):

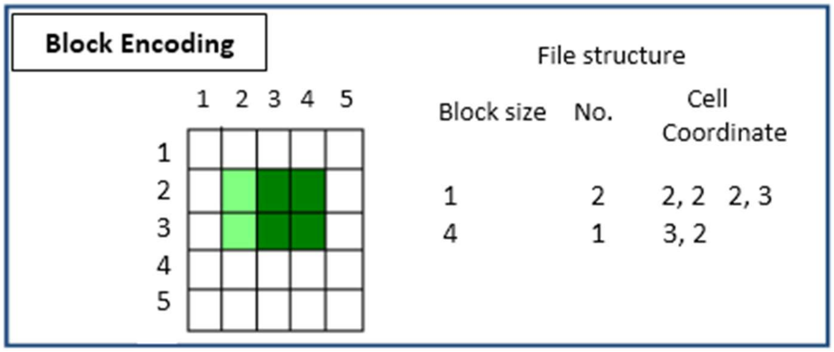
- Reduction of data on a row by row basis
- Stores a single value for a group of cells rather than storing values for individual cells.

- First line represents the dimension of the matrix (5×5) and the number of entities (1) present. In second and subsequent lines, the first number in the pair represents absence (0) or presence (1) of the entity and the second number indicates the number of cells referenced.



Block Encoding:

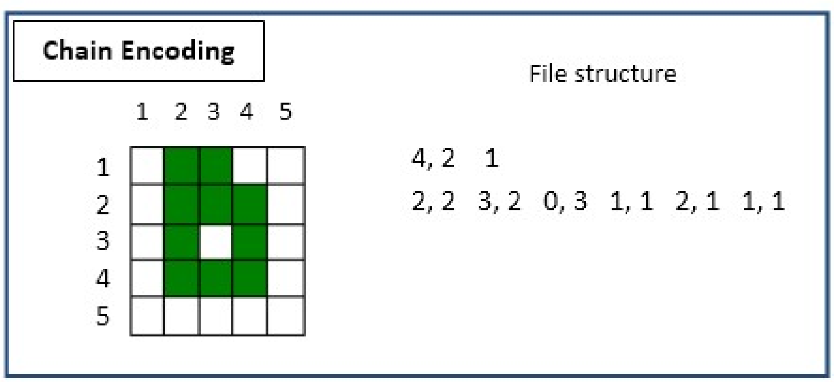
- Data is stored in blocks in the raster matrix.
- The entity is subdivided into hierarchical blocks and the blocks are located using coordinates.
- The first cell at top left hand is used as the origin for locating the blocks.
- Block encoding uses a technique called Medial Axis Transformation (MAT) to create the data structure. In this technique, the larger the square that may be fitted in any given region and the simpler the boundary, the more efficient compaction becomes.
- Both RLE and MAT are most efficient for large simple shapes and less so for small complicated areas.



Chain Encoding:

- Chain codes represent the raster boundary of a region by giving a starting point and the cardinal direction (east, north, west and south) to follow as we progress around the boundary.

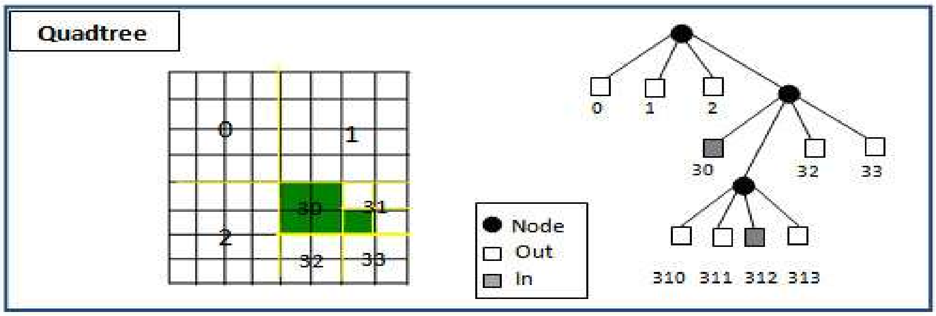
- Works by defining boundary of the entity i.e. sequence of cells starting from and returning to the given origin.
- Direction of travel is specified using numbers. (0 = North, 1 = East, 2 = South, 3 = West).
- The first line tells that the coding started at cell (4, 2) and there is only one chain. In the second line the first number in the pair tells the direction and the second number represents the number of cells lying in this direction.



Quadtree and Binary-tree Encoding:

- One problem with regular raster grids is that the resolution of the data limited by the size of the basic grid cells. The quadtree and binary-tree encoding provide an approach to addressing successively finer levels of details, with in principle, an infinite set of levels.
- The most efficient method of compact representation of space is based on successive, hierarchical division of a 2^n X 2^n array.

- If the division occurs by dividing the area into half each time, the method is known as binary-tree; if the region is tiled by subdividing the area step by step into quadrants, it is known as quadtree. In both the cases, the lowest limit of division is the single cell.
- A quadrant that cannot sub-divided is called a leaf-node.
- It is important to realize that quadtree is more concerned with the overall data model, rather than just saving space.



VECTOR DATA STRUCTURE

- The vector model is close to the traditional mapping approach where the objects are represented as points, lines, or areas. In a vector model, the positions of points, lines and areas are precisely specified.
- The position of each object is defined by a coordinate pairs. Vectors are graphical objects that have geometrical primitives such as points, lines, and polygons to represent geographical entities in computer graphics.
- Vectors have a precise direction, length and shape can be defined by coordinate geometry. A point is described by a single x-y co-ordinate pair and by its name or label. A line is described by a set of co-ordinate pairs and by its name or label.
- In reality, a line is described by an infinite number of points. In practice, this is not a feasible way of storing a line. Therefore, a line is built up by its starting and ending coordinate pairs.

OBJECT-BASED VECTOR MODEL

- The vector model is ideal to represent discrete entities. Discrete entities are represented as points, lines, and areas.

POINT:

- A location depicted by a single set of (x, y) coordinates at the scale of abstraction.

- The wells in a village, electricity poles in a town and cities in the world map are the examples of spatial features described by points.

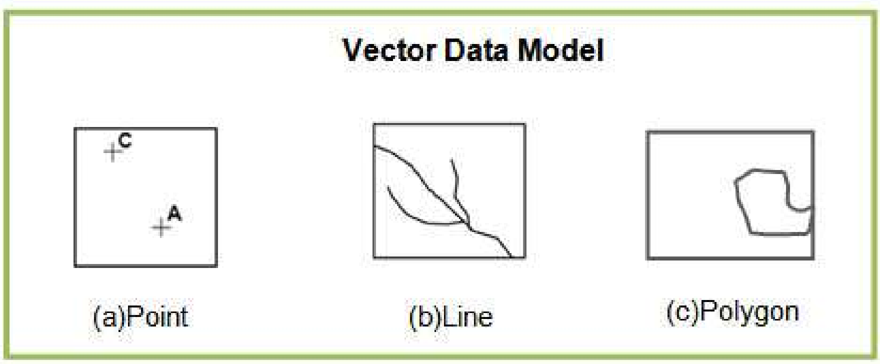
  Note: A city can be marked as a single point on a world map but would be marked as a polygon on a state map. The scale plays an important role in deciding the geometry of a geographical feature.

LINE/ARC:

- Ordered sets of (x, y) coordinate pairs arranged to form a linear feature. The curve in a linear feature is generated by increasing the density of points/ vertices.
- The roads, rails and telephone cables are examples of the spatial features described by lines.

POLYGON:

- The set of (x, y) co-ordinate pairs enclosing a homogenous area.
- The land parcels, agricultural farms and water bodies are the spatial features described by polygons.



- In the object-based vector model, points/lines/areas can be represented via different models. Four such models are as follows
  - Spaghetti Model
  - Vertex Dictionary Model.
  - Dual Independent Map Encoding (DMIE)
  - Topological Model.

Spaghetti Model:

- The spaghetti model uses the simplest type of data structure. All objects are defined as single items and no reference is made to other objects.

- Spaghetti data are collection of points and line segments with no real connection. There are no specific points that designate where the lines cross, nor there any logical relationships between the objects.
- The common boundaries between adjacent polygons are stored twice. This model cannot handle holes within a polygon. The structure is insufficient in data storage and queries, and consistency checks are not possible.

Vertex Dictionary Model:

- Vertex dictionary uses a similar approach as spaghetti but a smarter structure. It uses two files to store the vector data. The first file stores the vertices and the second file stores the description of objects as shown in figure.
- In this data model, if some vertices are shared by two adjacent polygons, these vertices are not required to store twice.
- However, topological relationships are not well defined in this model. The problem of island polygon still exists. Therefore, it is inferior for data analysis and query.

DIME Model:

- The DIME model was developed by the United States bureau of the census. This is a complex model but more intelligent. It uses three files to represent the vector data as shown in figure.
- This model also avoids duplication of data for adjacent polygons and can establish several relationships between objects.
- The DIME was the first attempt to built explicit topological relationships. However, it has several limitations compared to fully topographical model.

Topological Model:

- The mathematical field of topology investigates characteristics of geometry that remain invariant under certain transformations such as stretching or bending.
- The connections and relationships between objects are described independently of their coordinates.
- Topology in GIS is a set of objects and object data that defines the relationship between the objects. Topology refers to the relationships between spatial objects.
- Topology, as it relates to spatial data, consists of three elements such as adjacency, containment and connectivity.
- **Adjacency** and **Containment** describe the geometric relationships which exist between area features. Areas can be described as being adjacent when they share a common boundary.

- The topological model uses arc (chain) and nodes. Arc is the basic logical entity in this data model. Arc is a series of point that starts and end at a node. Because of this reason, topological model is often called Arc-node model.
- The advantage of topological data model are to avoid duplication in storing common boundaries of two polygons, and to solve problems when two versions of the common boundary do not coincide
- The disadvantage are to have to build very correct topological data sets without any single error. Topology can detect lines that do not meet correctly, polygons that are not closed properly, node or line segment.
- In practical applications of GIS, all possible relationships in spatial data should be used logically with more complicated data structures.
- The following topological relationships are commonly used are:
    - Point – point relationship.
    - Point – line relationship.
    - Point – area relationships.
    - Line – line relationships.
    - Line – area relationship.
    - Area – area relationship.

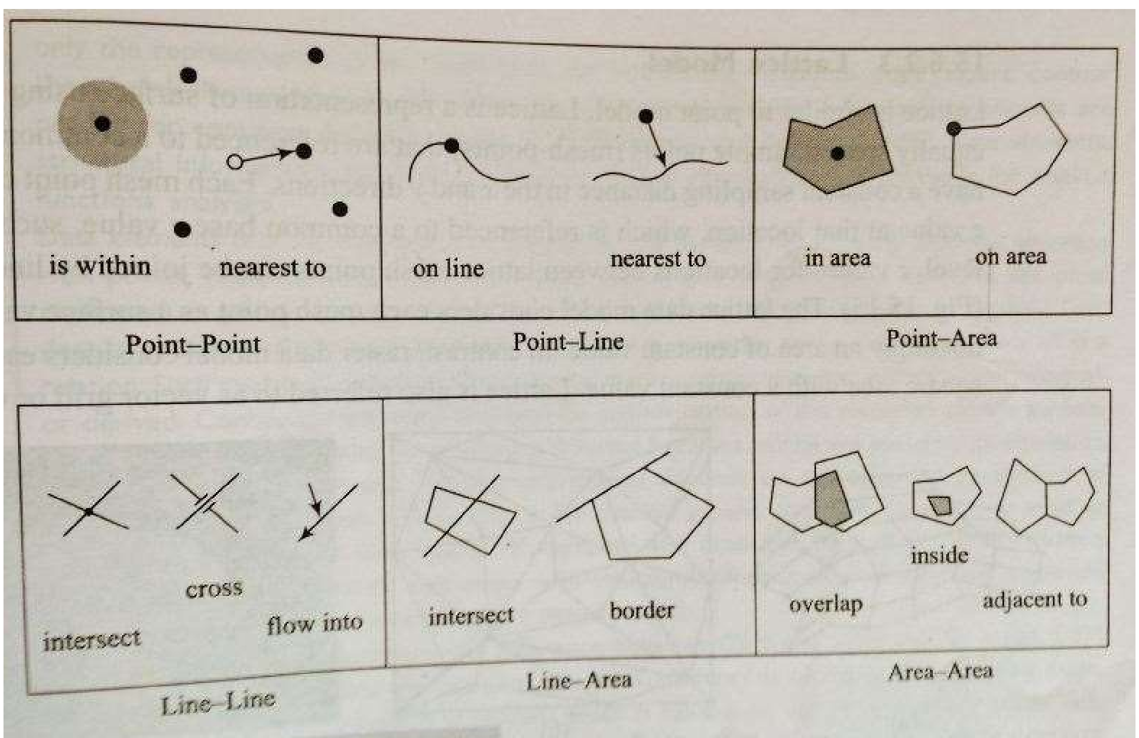


Fig. 15.14 Topological relationships

Field - Based Vector Model

- Although vectors are ideal for representing discrete objects, it can also be used for field-based or continuous data such as elevation, temperature etc.
- Mass points, contour lines/ isolines, lattice, and triangulated irregular network (TIN) are used to represent elevation or other continuously changing values.

Point Model:

- Multiple points are used to represent the surface. Mass point is a technique to represent surfaces using several points in a very dense manner.
- But point model is not a good approach to represent surfaces, because it requires the viewer to imagine the surface joining the points.

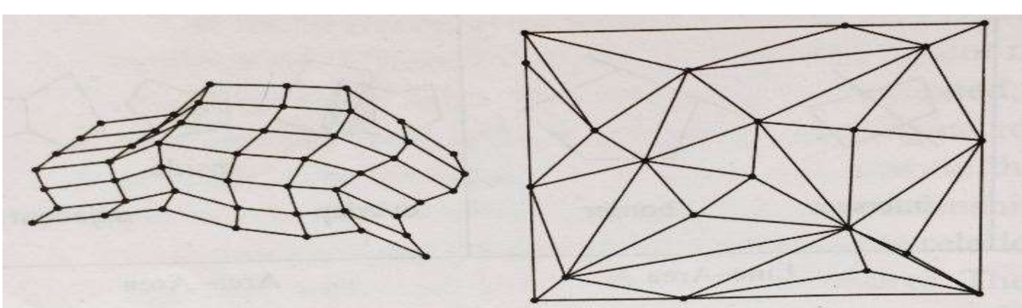
Contour/ Isolines Model:

- Contour is an imaginary line of constant elevation on the ground surface. The corresponding line on a map is called a **contour line.**
- A line on a map that joins places of the same elevation above sea level. Contour interval is the differences in elevation between two contour lines.

- Isolines is a line on a surface, connecting points of equal value such as temperature, rainfall etc.
- Contour/isolines can be used to represent surfaces.

Lattice Model:

- Lattice is similar to point model, it is a representation of surface using an array of equally spaced sample points(mesh-points) that are referenced to a common origin and have a constant sampling distance in the x and y directions
- Each mesh point contains the z value at that location, which is referenced to a common base z value, such as the sea level.
- The lattice model considers each mesh point as a surface value; it does not imply an area of constant value.
- In contrast, raster data model considers each cell as a square area with a constant value. Lattice is also referred to as vector grid or wire mesh.



(a) (b)

TIN Model:

- TIN or Triangulated Irregular Network stores GIS data for 3D surface model. In vector GIS, a TIN is used to create a DTM from either regular or irregular height data.
- The TIN method joins the height observations together with straight lines to create a non-overlapping mosaic of irregular triangles.
- A triangle consists of three lines connecting three identified with the coordinates of the three points forming it.
- The surfaces of individual triangle provide area, gradient(slope), and orientation (aspect). Gradient is the steepness of a unit of terrain, usually measured as an angle in degrees or as a percentage.
- Orientation is the direction in which a unit of terrain(triangles) faces, usually expressed in degree from north.

ADVANTAGES:

	Raster Model	Vector Model
1.	Simple data structure	It is a smaller file size
2.	Easy and efficient overlaying	Individual identity for discrete objects like line, polygon etc.
3.	Compatible with remote sensing imagery	Efficient for topological relationships.
4.	High spatial variability is efficiently represented.	Efficient projection transformation
5.	Efficient to represent continuous data.	Accurate map output and easy to edit.

Disadvantages:

	Raster Model	Vector Model
1.	Larger file size	Complex data structure

2.	All the objects are series of pixels, no identity for discrete objects other than points/pixels.	Difficult overlay operations.
3.	Difficult to build topological relationship.	High spatial variability is inefficiently represented.
4.	Inefficient projection transformation.	Not compatible with remote sensing imagery.
5.	Loss of information when using larger cells	Not appropriate to represent continuous data
6.	Difficult to edit.	

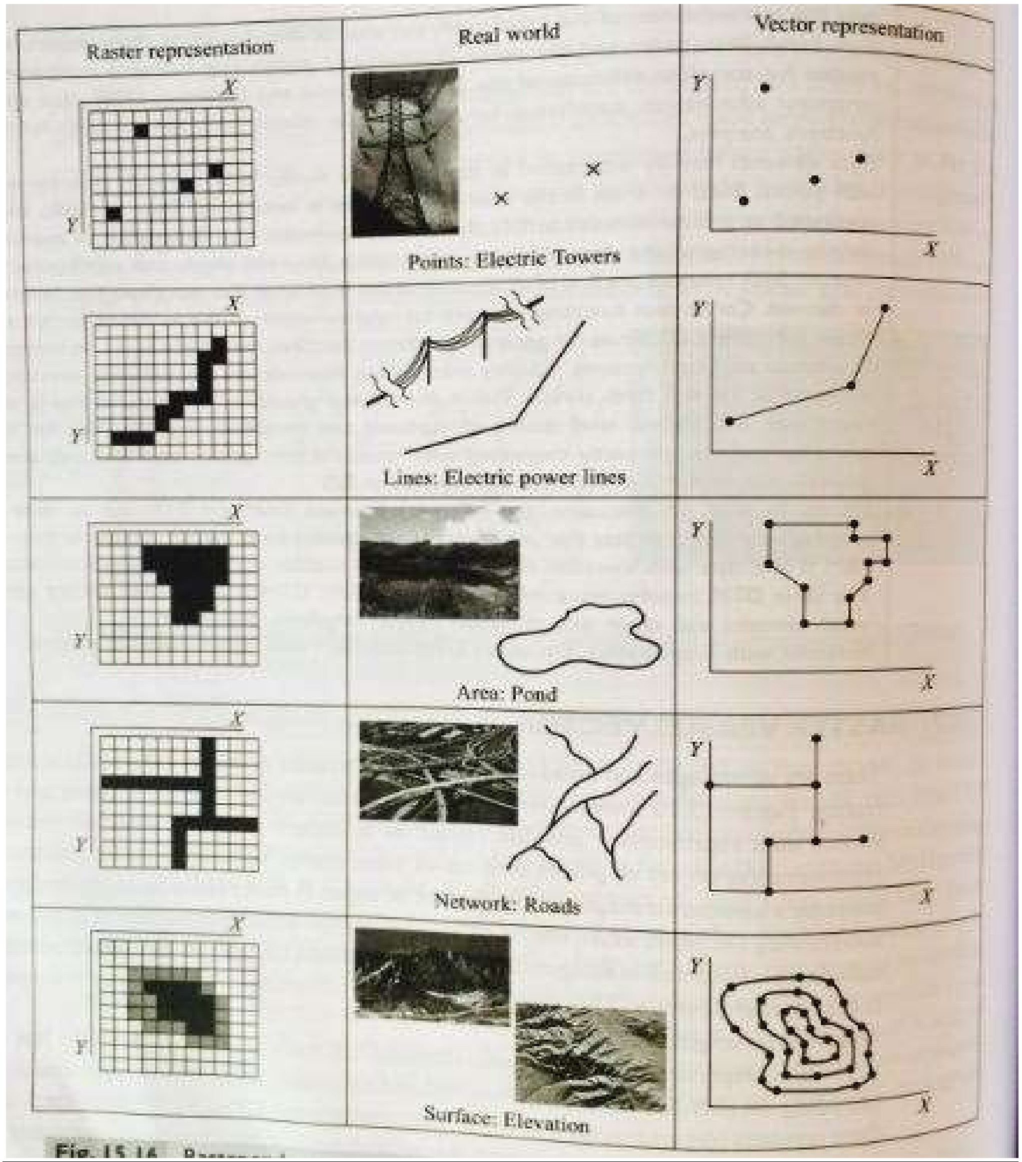


Fig- RaSter and Vector Representation

FUNCTIONING OF GIS:

The functional requirements of GIS are:

- If we could relate information about the rainfall to remote sensing imageries of our country, we might be able to tell which wetlands dry up at certain times of the year.
- The primary requirement for the source of data is that locations for the variables are known, location may be annotated by x, y and z coordinates of longitude, latitude and elevation.
- A GIS can also convert existing digital information, which may not yet be in map form, into forms it can recognize and use.

Data Capture:

- If the data to be used are not already in digital form, various techniques can capture the information.
- Maps can be digitalized with a digitizer table or hand-traced with a computer mouse to collect the coordinates of features.
- Electronic scanning devices will also convert map features into digital format.

Database Storage and Management:

- After data are collected and integrated, the GIS must provide facilities which can contain and maintain data.
- Effective data management has many definitions but should include all of the following aspects: data security, data integrity, data storage and retrieval and data maintenance abilities.

Data Integration:

- A GIS makes it possible to link, or integrate information that is difficult to associate through any other means. Thus a GIS can use combinations of mapped variables to build and analyze new variables.
- For example by using GIS technology and water-company billing information, it is possible to stimulate the discharge of materials in the septic systems in a neighborhood upstream from a wetland.
- The bills show how much of water is used at each address. The amount of water used by a customer roughly predicts the amount of material that is discharged into the septic systems, so that areas of heavy septic discharge can be located using a GIS.

Data Base Management System:

- Generally, all data have three major components

- Position(spatial data)
- Attributes (thematic),
- Time (temporal)

- In GIS, attribute data are linked with spatial data or geometric data. Some GIS software can handle attribute data very efficiently, some are not capable of handling attribute data, and rely on links with conventional database management system(DBMS) software and others have very limited database handling capabilities though they possess efficient analysis facilities.
- **Database** is a collection of records stored in a computer in a systematic way, so that it is consulted by a computer program to answer questions. For better storing and retrieval, each record is usually organized as a set of related data elements. The computer program used to manage and query a database is known as **DBMS (**Database management systems).

FUNCTIONS OF DBMS:

The following are the functions performed by a typical DBMS.

- The DBMS provides functions to define the structure of the data in the application. These include defining and modifying the record structure, the type and size of fields, and the various conditions to be satisfied by the data in each field.
- Data manipulation- the functions which perform these operations are also part of the DBMS. These functions can handle planned and unplanned data manipulation needs.
- Data security and integrity- the DBMS contains functions that handle the security and integrity of data in the application. These can be easily invoked by the application and hence the application programmer need not code these functions in her programmers.
- Data recovery and concurrency- Recovery of data after a system failure and concurrent access of record by multiple users are also handles by the DBMS.
- Data dictionary maintenance- Maintaining the data dictionary, which contains the data definition of the application, is also one of the functions of a DBMS.
- Performance- Optimizing the performance of the queries is one of the important functions of a DBMS. Thus the DBMS provides an environment that is both convenient and efficient to use when there is a large volume of data and many transactions to be processed.

APPLICATIONS OF GIS:

GIS is involved in various areas. These include topographical mapping, socioeconomic and environment modeling, and education. The role of GIS is best illustrated with respect to some of the representative application areas that are mentioned below:

Tax Mapping:

- Raising revenue from property taxes is one of the important functions of the government agencies. The amount of tax payable depends on the value of the land and the property.
- The correct assessment of value of land and property determines the equitable distribution of the community tax.
- A tax assessor has to evaluate new properties and respond to the existing property valuation. To evaluate taxes the assessor uses details on current market rents, sale, maintenance, insurance and other expenses.
- Managing as well as analyzing all this information simultaneously is time consuming and hence comes the need of GIS. Information about property with its geographical location and boundary is managed by GIS.
- Querying the GIS database can locate similar type of properties in an area. The characteristics of these properties can then be compared and valuation can be easily done .

Business:

- Approximately 80 percent of all business data are related to location. Businesses manage a world of information about sales, customers, inventory, demographic profiles etc.
- Demographic analysis is the basis for many other business functions: customer service, site analysis, and marketing. Understanding your customers and their socioeconomic and purchasing behavior is essential for making good business decisions.
- A GIS with relevant data such as number of consumers, brands and sites they go for shopping can give any business unit a fair idea whether their unit if set up is going to work at a particular location the way they want it to run.

Logistics:

- Logistics is a field that takes care of transporting goods from one place to another and finally delivering them to their destinations.
- It is necessary for the shipping companies to know where their warehouses should be located, which routes should the transport follow that ensures minimum time and expenditures to deliver the parcels to their destinations. All such logistics decisions need GIS support.

Environment:

- GIS is being increasingly involved in mapping the habitat loss, urban sprawl, land-use change etc.

- Mapping such phenomena need historical land use data, anthropogenic effects which greatly affect these phenomena are also brought into GIS domain.
- GIS models are then run to make predictions for the future.

Emergency evacuation:

- The occurrence of disasters is unpredictable. We as humans are unable to tell when, where and what magnitude of disaster is going to emerge and therefore solely depend on disaster preparedness as safety measures.
- It is important to know in which area the risk is higher, the number of individuals inhabiting that place, the routes by which the vehicles would move to help in evacuating the individuals.
- Thus preparing an evacuation plan needs GIS implementation.

Remote sensing applications in Water Shed Management:

- Scientific planning and management is essential for the conservation of land and water resources for optimum productivity.
- Watersheds being the natural hydrologic units, such studies are generally carried out at watershed scale and are broadly referred under the term watershed management.
- It involves assessment of current resources status, complex modeling to assess the relationship between various hydrologic components, planning and implementation of land and water conservation measures etc.
- **Water resource mapping, land cover classification, estimation of water yield and soil erosion, estimation of physiographic parameters** for land prioritization and water harvesting are a few areas where remote sensing techniques have been used.
- The remote sensing applications in water resources management under the following five classes:

- Water resources mapping
- Estimation of watershed physiographic parameters
- Estimation of hydrological and meteorological variables
- Watershed prioritization
- Water conservation