LECTURE NOTES ON
REMOTE SENSING & GIS
IV B. Tech II semester (JNTU (A)-R13)

C.THEJASWINI VEENA
Assistant Professor

G.PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY

CIVIL ENGINEERING

OBJECTIVES:

1. *To provide an exposure to GIS and its practical applications in civil engineering*
2. *Analyze the energy interactions in the atmosphere and earth surface features*

OUTCOMES:

On completion of the course the students will have knowledge on

1. *Principles of Remote Sensing and GIS*
2. *Analysis of RS and GIS data and interpreting the data for modeling applications*

UNIT – III

GEOGRAPHIC INFORMATION SYSTEM:

Introduction, GIS definition and terminology, GIS categories, components of GIS, fundamental operations of GIS, A theoretical framework for GIS.

TYPES OF DATA REPRESENTATION:

Data collection and input overview, data input and output. Keyboard entry and coordinate geometry procedure, manual digitizing and scanning, Raster GIS, Vector GIS – File management, Spatial data – Layer based GIS, Feature based GIS mapping.

TEXT BOOKS:

1 Remote Sensing and GIS by B.Bhatta, Oxford University Press,New Delhi.
2 Fundamentals of remote sensing by gorge Joseph , Universities press, Hyderabad

REFERENCES:

1. Advanced surveying : Total station GIS and remote sensing – Satheesh Gopi – Pearson publication.
2. Remote Sensing and its applications by LRA Narayana University Press 1999.
3. Basics of Remote sensing & GIS by S.Kumar, Laxmi Publications.
4. Remote sensing and GIS by M.Anji Reddy ,B.S.Pubiliications,New Delhi.
5. GIS by Kang – tsung chang, TMH Publications & Co.,

UNIT-3
A GEOGRAPHIC INFORMATION SYSTEM (GIS)

Introduction and Definition

Geographic Information System (GIS) is becoming more & more popular among decision makers as it enables them to quickly refer the GIS outputs which help them in solving problems and making right decisions. Visualization of features, converting data into need-based maps (thematic maps) and capability of providing solutions by taking into account overall scenario of an area are some of the virtues of GIS due to which it is being implemented across a number of sectors and departments (e.g.,transportation, forestry, environment, disaster management, urban planning, health etc).

We often feel difficulty to understand and visualize a problem just by seeing data or figures. We are more comfortable with visual representation of a problem that's what GIS is giving to us (that too with the true representation of real-world).

Definition

GIS is made up of three terms- Geographic, Information and System. In literal meaning Geographic Information System is a System containing Information which is geographic in nature.

GIS can be defined as - A System which involves collecting/capturing, storing, processing. manipulating, analyzing, managing, retrieving and displaying data (information) which is, essentially, referenced to the real-world or the earth (i.e. geographically referenced). **Explanation of the Definition**We have used many term in the definition of GIS mentioned above. It is necessary to discuss each term for getting an idea - what actually GIS is?

Collection/Capturing

The dataset collected for GIS may be in the form of hard copy maps, satellite images, survey data or other data obtained from other primary and secondary sources. Collection of data depends on the objective of the assignment. Data capturing involves digitization of hard copy maps and satellite images.

Storage

In GIS Storage means not merely storing whatever data we have collected. The collected data is converted in usable GIS format and then finally stored for further use either on computer hard disk or in other storage devices (CD, DVD, magnetic tapes etc.)

Processing and Manipulation

The collected and stored dataset is imported and converted into layers. Then required attributes are attached. Then data is processed for refinement, removing errors and preparing it for further GIS-based analysis. Data manipulation is essential so that it can be represented in proper understandable form.

Analysis

Analysis of GIS data is required to convert it into desired outputs. There are many type of analysis in GIS which is (or are) to be done is objective dependant. The analysis may be statistical, spatial or specialized (like network analysis, utility analysis etc. Need not to say GIS analysis requires skilled professionals.

Management

Data management is essential and very important part of GIS for storing, managing and properly maintaining GIS database.

Retrieval

In GIS, data can be retrieved through SQL or spatial queries. Some software provide tools to retrieve data by simply selecting the features. Retrieval is used for getting information about the features of our interest.

Display

Displaying of final output may in many forms. These may be hard copy printouts, on-screen display of maps, internet-based map display (through Internet Map Servers) or in the form of presentation (like power point).

GIS and Remote Sensing

a. **GIS in remote sensing**

For the users of remote sensing, it is not sufficient to display only the results obtained from image processing. For example, to detect land cover change in an area is not enough, because the final goal would be to analyse the cause of change or to evaluate the impact of change. Therefore the result should be overlaid on maps of transportation facilities and land use zoning as shown in Figure 13.1.1. In addition, the classification of remote sensing imagery will become more accurate if the auxiliary data contained in maps are combined with the image data.

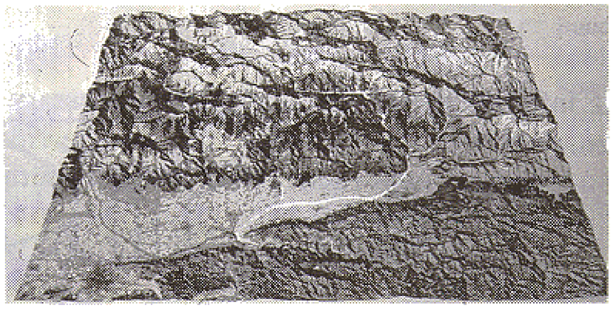

Figure 13.1.1 An overlay of RS data and map data (elevation data)

Interpretation can be made easier by an overlay of geo-corrected RS data with a map data

In order to promote the integration of remote sensing and geographic data, **geographic information system (GIS)** should be established in which both the image and graphic data are stored in a digital form, retrieved conditionally, overlaid on each other and evaluated with the use of a model.Figure 13.1.2 shows a comparison between the computer assisted GIS and the conventional analog use of maps.

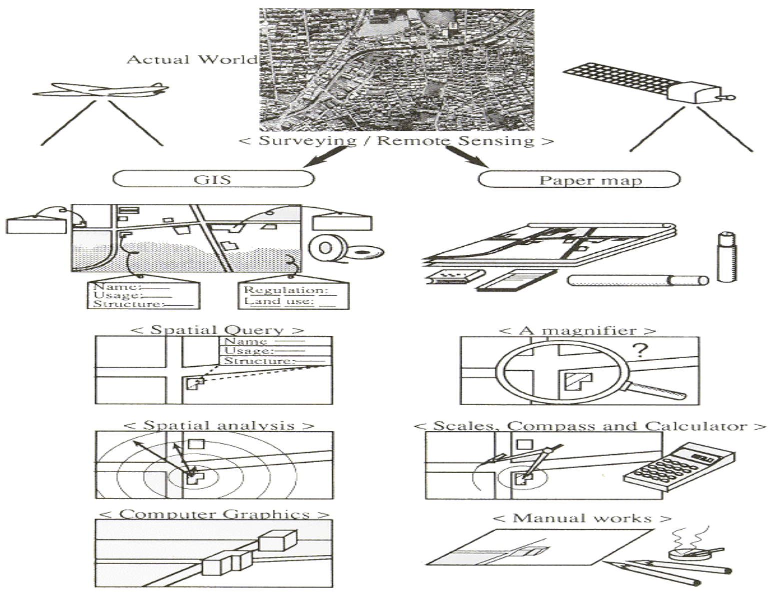

Figure 13.1.2 Geo-information management with GIS
- A comparison with paper maps -

b. Function of GIS

The following three functions are very important in GIS.

(1) To store and manage geographic information comprehensively and effectively

(2) To display geographic information depending on the purpose of use

(3) To execute query, analysis and evaluation of geographic information effectively

At present, the following research and development have been undertaken.

(1) Model and data structure for GIS

(2) Data input and edition

(3) Spatial query

(4) Spatial analysis

(5) Visualization

BASIC CONCEPT OF GIS

Geographic relates to the surface of the earth.

Information is a knowledge derived from study, experience, or instruction.

System is a group of interacting, interrelated, or interdependent elements forming a complex whole.

Science is the observation, identification, description, experimental investigation, and theoretical explanation of phenomena.

Components of a GIS

GIS is an organized collection of computer hardware, software, geographic data, procedures, and personnel designed to handle all phases of geographic data capture, storage, analysis, query, display, and output.

A working GIS integrates five key components: hardware, software, data, people, and methods.

Hardware

Hardware is the computer on which a GIS operates. Today, GIS software runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations.

Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are

- Tools for the input and manipulation of geographic information
- A database management system (DBMS)
- Tools that support geographic query, analysis, and visualization
- A graphical user interface (GUI) for easy access to tools

Data

Possibly the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house or purchased from a commercial data provider. A GIS will integrate spatial data with other data resources and can even use a DBMS, used by most organizations to organize and maintain their data, to manage spatial data.

People

GIS technology is of limited value without the people who manage the system and develop plans for applying it to real-world problems. GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work.

Methods

A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

Functional operations of GIS:

1. Data Input
2. Existing maps, field observations, aerial photographs, sensors (air boarn & satellites)
3. Data storage & management
4. Data processing & analysis
5. Data output & presentation: Maps/tables/figures/video/audiol3d models.
6. Various interaction techniques: mouse ,key board ,voice etc

TYPES OF DATA REPRESENTATION:

GIS data can be broadly described as- Spatial data and Non-spatial data.

SPATIAL DATA

Spatial data is geographical representation of features. In other words, spatial data is what we actually see in the form of maps (containing real-world features) on a computer screen. Spatial data can further be divided into two types- **vector**and **raster data**.

Vector Data

Vector data represents any geographical feature through point, line or polygon or combination of these.

1. Point

A point in GIS is represented by one pair of coordinates (x & y). It is considered as dimension-less object. Most of the times a point represent location of a feature (like cities, wells, villages etc.).

2. Line

A line or arc contains at least two pairs of coordinates (say- x1, y1 & x2, y2). In other words a line should connect minimum two points. Start and end points of a line are referred as nodes while points on curves are referred as vertices. Points at intersections are also called as nodes. Roads, railway tracks, streams etc. are generally represented by line.

3. Polygon

In simple terms, polygon is a closed line with area. It takes minimum three pairs of coordinates to represent an area or polygon. Extent of cities, forests, land use etc. is represented by polygon.

Raster Data

Raster data is made up of pixels. It is an array of grid cells with columns and rows. Each and every geographical feature is represented only through pixels in raster data. There is nothing like

point, line or polygon. If it is a point, in raster data it will be a single pixel, a line will be represented as linear arrangement of pixels and an area or polygon will be represented by contiguous neighbouring pixels with similar values.
In raster data one pixel contain only one value (unlike vector data where a point, a line or a polygon may have number of values or attributes) that's why only one geographical feature can be represented by a single set of pixels or grid cells. Hence a number of raster layers are required if multiple features are to be considered (For example- land use, soil type, forest density, topography etc.).

As discussed earlier digital satellite images are also in raster format.

NON-SPATIAL DATA

Attributes attached to spatial data are referred to as non-spatial data. Whatever spatial data we see in the form of a colourful map on a computer screen is a presentation of information which remains stored in the form attribute tables. Attributes of spatial data must contain unique identifier for each object. There may be other field also containing properties/information related a spatial feature. Attribute table of spatial data also contains 'x' and 'y' location (i.e. latitude/longitude or easting/northing) of features; however in some GIS software these columns may remain 'invisible'.

For example- if we are doing demographic analysis of villages then attributes of each point (representing a village) must have a unique village ID and other demographic information like total population, number of males & females, number of children etc.
In another example- if we are doing some GIS analysis related to road then each road must have its unique Road ID. Other attributes may include like road length, road width, current traffic volume, number of stations etc.

Vector Data

- Represented by point, line and polygon.
- Relatively small file size (small data volume)
- Excellent representation of networks.
- A large no. of attributes can be attached, hence more information intensive and a number of thematic maps can be prepared from a single layer.
- Features are more detailed & accurate.
- Creating, cleaning and updating data is more time and labour consuming.
- Topology-based analysis & operations are easier to perform (like network analysis etc.).
- Can not represent continuous values like land use, elevation etc very well.
- Assigning projection and transformations are less time taking and consumes less memory of the computer system.
- Topology makes data structure complex.

Raster Data

- Points, line & polygons everything in the form of Pixels.
- Large file size.
- Networks are not so well represented.

- Only one pixel value represents each grid cell.
- Generalization of features (like boundaries) hence accuracy may decrease.
- Simulations and modeling is easier (spatial analysis, terrain modeling etc.).
- Maintaining is easier.
- Excellent for representing data containing continuous values (like land use, elevation etc.)
- Coordinate-system transformations take more time and consume a lot of memory.
- Grid cells or pixel makes simpler data structure.

Human resources in GIS:

People in GIS are one of the five essential components. Success of a GIS assignment largely depends upon the human resources involved in it. A good, skilled, motivated and mutually cooperating team produces excellent results. Like other fields, in GIS also there is a chain of tasks interlinked with each other; any weak link may lead to hampering the whole assignment.

GIS Manager

A GIS manager is the most important personnel in GIS team. Here the term GIS manager is used in broader perspective (it covers GIS team leader also). His responsibilities are to conceptualize, manage and effectively implement GIS in a project. A good GIS manager always explores the possibilities of applying GIS in relevant projects for saving time, better presentation and effective solutions. He should have skills & experience (at least 5-7 years) in variety of GIS tasks and good management skills. A GIS manager should always motivate and “push” the team to finish tasks in time with quality. He should be able to conceptualize the whole GIS task with precise methodology and should also have back up plans, in case if one fails can immediately implement the other. Most of the time GIS-based tasks in a project are critical ones as the team has to bring out good output from bad quality and deficient data, hence GIS manager should have clear picture of the whole assignment and should implement right methods (considering time constraint!).

GIS Specialist

A GIS specialist (or GIS engineer) should have basic as well as advance skills in GIS. He should have at least three years of experience in various applications of GIS (like- transportation, water resource, disaster management, urban planning etc.). He should possess problem defining and solving abilities. A GIS specialist is expected to support GIS manager in defining methodology and helping in its proper implementation. Possessing knowledge of basics of computer programming and database always advantageous for GIS specialist (and for organization too).

GIS Programmer

GIS programmers are now-a-days high in demand. They customize GIS software and develop need-specific GIS modules with in given constraints (time & budget). They should have sound knowledge of computer programming (Visual Basic, C++ and JAVA are most frequently used in GIS development) and should be able to understand GIS problems (hence basic GIS knowledge is required). Experienced GIS programmers are often expected play role of database managers/administrators; hence they should have expertise in database management (in popular database software like- Oracle, MySQL etc.).

GIS Support Staff

GIS support staff performs GIS-related tasks at basic level which includes data collection, field survey, map scanning, data entry, digitization, georeferencing, map printing etc. GIS support staff should have basic knowledge of GIS. It is very important for GIS managers and specialists to convey the support staff that how important there inputs are and doing mistakes at their level may lead to disasters at higher level (while doing analysis and modeling). The actual of GIS begins at this level hence support staff should avoid as much errors as possible in their tasks.

Data Input and Editing

Role of data input and editing

Data acquisition occupies about 80 percent of the total expenditure in GIS. Therefore data input and editing are very important procedures for the use of GIS.

Initial data input

Geometric data as well as attribute data are input by the following methods.

(1)Direct data acquisition by land surveying or remote sensing
Vector data can be measured with digital survey equipment such as total stations or analytical photogrammetric plotters. Raster data are sometime obtained from remote sensing data.

(2)Digitization of existing maps (see Figure 13.3.1)

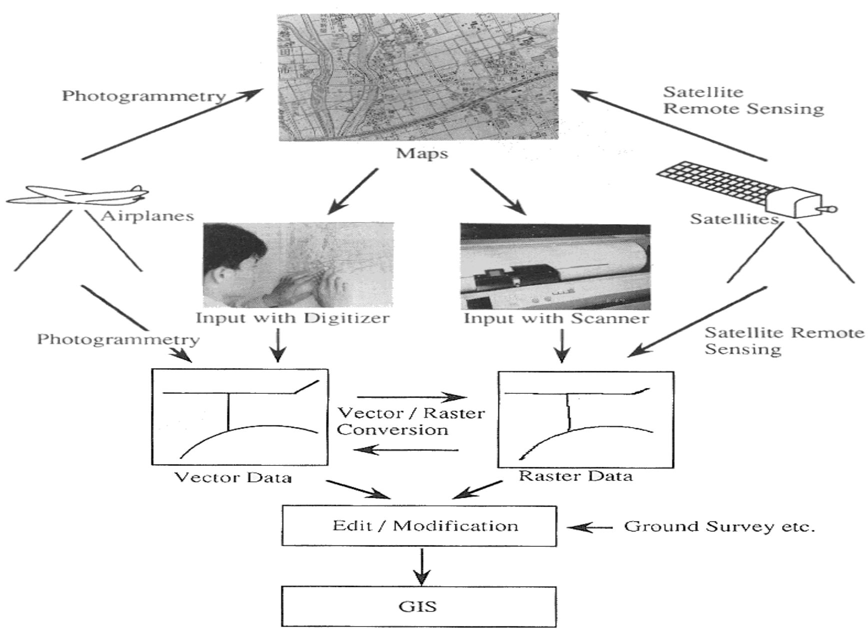

Figure 13.3.1 A process of input / update of geographic data

Existing maps can be digitized with a scanner or tablet digitizer. Raster data are obtained from a scanner while vector data are measured by a digitizer. In GIS, raster data and vector data are frequently converted to vector data and raster data respectively, which are called raster/vector conversion and vector/raster conversion respectively.

Editing

Editing is needed to correct, supplement and add to the initial input data through interactive communication on a graphic display using the following procedures.
(1) To input manually or interactively those complicated attributes which are not effectively digitized in the initial input stage.
(2) To correct errors of input data or to supplement with other data.

Problems in Data Input and editing

there are two main problems.

(1)Manual operations
It is difficult to automate data input and editing because of unremovable noise and incomplete original maps, which result in a large amount of manual work with resultant inefficiencies in time and cost.
(2) Unreliability of input data

As the input involve many kinds of errors, mistakes and misregistration because of the manual input, further effort should be applied to obtain data high quality and reliability.

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: 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. (Discussed in detail in Ch. 5)

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.
- There are five types of data entry systems commonly used in a GIS:
  - **keyboard entry**
  - **coordinate geometry**
  - **manual digitizing**
  - **scanning**
  - **input of existing digital files**

Keyboard entry: involves manually entering the data at a computer terminal. Attribute data are commonly input by keyboard whereas spatial data are rarely input this way.
Keyboard entry may also be used during manual digitizing to enter the attribute information. However this is usually more efficiently handled as a separate operation.
Roads files versus the census file -- roads file will use codes for the various road types while the census file uses exact numbers for things like total population, age range, etc.

Coordinate Geometry (COGO): involves entering survey data using a keyboard. From these data the coordinate of spatial features are calculated. This produces a very high level of precision and accuracy which is needed in a cadastral system.

For a city with 100,000 parcels, it would cost approximately $1 - $1.50 per parcel or $100,000 to $150,000 to digitize the parcels manually. COGO procedures are commonly 6 times and can be up to 20 times more expensive than manual digitizing.
Surveyors and engineers want the higher accuracy of COGO for their applications. Planners and most others are happy with the lower accuracy provided by manual digitizing.
Manual Digitizing: The most widely used method for entering spatial data from maps. The map is mounted on a *digitizing tablet* and a hand held device termed a puck or cursor is used to trace each map feature. The position of the puck is accurately measured by the device to generate the coordinate data.

- Digitizing surfaces range from 12 inches x 12 inches (digitizing tablet) to 36 x 48 (digitizing table) and on up.
- The digitizing table electronically encodes the position of the pointing device with a precision of a fraction of a millimeter.
- The most common digitizer uses a fine wire mesh grid embedded in the table. The cursor normally has 16 or more buttons that are used to operate the data entry and to enter attribute data.
- The digitizing operation itself requires little computing power and so can be done without using the full GIS. A smaller, less expensive computer can be used to control the

digitizing process and store the data. The data can later be transferred to the GIS for processing. The problem with this is having enough software for all the computers.

- The efficiency of digitizing depends on the quality of the digitizing software and the skill of the operator. The process of tracing lines is time-consuming and error prone. The software can provide aids that substantially reduce the effort of detecting and correcting errors.
- Attribute information may be entered during the digitizing process, but usually only as an identification number. The attribute information referenced to the same ID number is entered separately.
- Manual digitizing is a tedious job. Operator fatigue (eye strain, back soreness, etc.) can seriously degrade the data quality. Managers must limit the number of hours an operator works at one time. A commonly used quality check is to produce a verification plot of the digitized data that is visually compared with the map from which the data were originally digitized.

Scanning: Scanning provides a faster means of data entry compared to manual digitizing.
In scanning, a digital image of the map is produced by moving an electronic detector across the surface of the map.

There are two types of scanner designs:

Flat-bed scanner**:** On a flat-bed scanner the map is placed on a flat scanning stage and the detectors move across the map in both the X and the Y directions (similar to copy machine).

Drum scanner**:** On a drum scanner, the map is mounted on a cylindrical drum which rotates while the detector moves horizontally across the map. The sensor motion provides movement in the X direction while the drum rotation provides movement in the Y direction.
The output from the scanner is a digital image. Usually the image is black and white but scanners can record color by scanning the same document three times using red, green and blue filters.

Inputting existing digital files: There are many companies and organizations on the market that provide or sell digital data files often in a format that can be read directly into a GIS. These digital data sets are priced at a fraction of the cost of digitizing existing maps.
Over the next decade, the increased availability of data should reduce the current high cost and lengthy production times needed to develop digital geographic data bases.

SCANNING VERSUS MANUAL DIGITIZING

Scanning is being used by many organizations, yet the subject is very controversial. One reason for the questions on data accuracy is that rigorous trials are few and of necessity are specific to the organization and application.
Data entry using scanning is claimed to be 5 to 10 times (or more) faster than digitizing.
However maps normally must be redrafted before they can be scanned or the color separates must be scanned.
Redrafting is often considered to be a major disadvantage of the scanning option. Redrafting, although time consuming, does not necessarily add to the cost of the data conversion process.

Redrafting can reduce the total cost of both scanning and manual digitizing. For example, studies by the US Forest Service have shown that a "map preparation" step before the manual digitizing is done can reduce the overall digital encoding costs by as much as 50%.

DIGITAL ELEVATION DATA

Digital elevation data are a set of elevation measurements for locations distributed over the land surface. They are used to analyze the topography (surface features) of an area.
Various terms have been used to refer to digital elevation data and its derivatives:
Digital Terrain Data Digital Terrain Models Digital Elevation Model
Digital Terrain Elevation Data
Digital elevation data are used in a wide range of engineering, planning, and military applications. For example, the are used to:

- Calculate cut-and-fill operations for road construction;
- Calculate the area that would be flooded by a hydroelectric dam;
- Analyze and delineate area that can be seen from a location in the terrain;
- Intervisibility can also be used to plan route locations for roadways;
- Optimize the location of radar antennas or microwave towers; or
- Define the viewshed of an area.

The methods used to capture and store elevation data can be grouped into four basic approaches:

* A regular grid contours profiles
* Triangulated Irregular Network (TIN) SEE FIGURE 4.9 page 122
* Digital elevation data are generated from existing contour maps, by photogrammetric analysis of stereo aerial photographs, or more recently by automated analysis of stereo satellite data.
* DTM data are most commonly provided in grid format in which an elevation value is stored for each of a set of regularly spaced ground positions. Each data point represents the elevation of the grid cell in which it is located.

One of the limitations of the raster form of representation is that the same density of elevation points is used for the entire coverage area.
Ideally, the data points would be more closely spaced in complex terrain and sparsely distributed over more level areas. A number of methods have been developed to provide a variable point density.

One method is to use a variable grid cell spacing to accommodate a variable density of points, with smaller cell sizes being used to capture the detail in more complex terrain.

A second approach has been to use irregularly spaced elevation points and represent the topography by a network of triangular facets. In this way, elevation data can be stored and manipulated using a vector representation. The **TIN** is produced from a set of irregularly spaced elevation points (SEE FIGURE 4.9). A network of triangular facets is fit to these points. The coordinate positions and elevations of the three points forming the vertices of each triangular facet are used to calculate such terrain parameters as the slope and aspect.

The advantage of a TIN compared with a gridded representation is that the TIN can use fewer points, capture the critical points that define discontinuities like ridge crests, and can be topologically encoded so that adjacency analyses are more easily done.

A third way to digitally represent a topographic surface is by development of a profile showing the elevation of points along a series of parallel lines. Elevation values should be recorded at all breaks in slope and at scattered points in level terrain. If the profiles are constructed from a topographic map, the elevation values can only be taken where the profile crosses a contour line.

The fourth approach is to digitize contour lines. Here the topographic surface is represented by series of elevation points taken along the individual contours. Although elevation data can be converted from one format to another, each time the data are converted some information is lost reducing the detail to the topographic surface.

- Digital elevation data is available in the US and was first produced by the Defense Mapping Agency. They were produced by scanning the contour overlays for 1:250,000 scale topographic maps.
- These data have an accuracy of 15 m in level terrain, 30m in moderate terrain, and 60 m in steep terrain.
- The data are sold by the map sheet as 1 degree x 1 degree blocks and are available for the entire US.
- The USGS plans to progressively upgrade the accuracy of this data set and is also producing a higher accuracy DTM file with a 30m sampling interval. The data are maintained in two datasets; one with a +7m accuracy and the other with a +7 - +15m accuracy. These data are available for about 30% of the US and are sold by 7.5 minute quad sheets.
- The unit price for these data decrease with the number of DTs purchased. Prices for orders of six or more DTM consist of a base charge of $90 and $7 for each additional unit.

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 he 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.

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. Spatial data is often accessed, manipulated or analyzed through Geographic Information Systems (GIS).