INTRODUCTION AND PRINCIPLES OF GIS

D. S. Rathore
National Institute of Hydrology, Roorkee

1.0 Introduction

A Geographic Information system is computer based system for input, storage, manipulation, output of geographic or spatial data and their attributes. Example of geographic data are physical, geological, hydrographic, soil, transportation, land use, geo morphological, utilities, cultural, historic sites/ tourism maps etc. For example the attribute information for the groundwater well geographic data are well identifier, water depths in various seasons and years, quality, draft, type, year of digging, ownership, uses and other descriptions. For river geographic data, the attribute data can be water quality, discharge and morphology, roughness, use, name, order etc. In a GIS the geographic and attributes data both are handled with ease.

In the past, the geographic information was handled manually. This caused the up keep of large sheets and data files tedious. The manual handling of the data also led to errors. The analysis of the two or more data/ maps required going through them to and fro and this led to physical and mental strain to an interpreter or analyst. Mathematical computations were also difficult. The use of GIS removed these hurdles in using data/ maps. Further, advantage of GIS is that the analysis that are computationally intensive can be easily handled. Many more analysis techniques have evolved.

Technological development e.g. faster computer, higher storage and high speed of data access etc. have made the use of already available analysis capabilities more effective. At the same time, the developments in the software also provided improved capabilities. The use of software became easy due to development in the user interface. Thus, the GIS have become more powerful, efficient and user friendly. This has led to new applications of this technique. The data can now be effectively manipulated. For planners, this tool is very useful as the geographic information is available at a click of the button to them.

2.0 Data Types

Point: Point data does not have length, width or area. Examples of point data are villages, cities, wells, springs etc. The points data in conventional maps are shown by symbols e.g. wells are shown as blue color filled circles, cities are shown by red filled square, springs are shown by blue color plus sign etc. Point data in one map may be shown as polygon data in another map e.g. cities are shown by many red unfilled squares in topographic maps.

Line: Line data do not have width but only length. They can be either straight lines or curved lines. Examples of line data are streams, canals, rails, roads, fire lines, telegraph lines, power lines, boundaries of land use classes, watershed boundaries, village boundaries etc.

Area: Area data are enclosed by the regular or irregular polygons. They have attributes e.g. area, shape, perimeter etc. Examples of area data are ponds, reservoirs, watersheds, forests etc.

3.0 Data Model and Data Structure

While data models describe the conceptualization of geographic data, the data structures describe how the data are actually stored and handled in the computer. They are also called high and low level representations of the geographic data respectively.

Raster: In raster data model, the geographic area is divided in grid cells (pixel or picture element). The information for grid cells are stored by the GIS. Location of the Grid cells is known to the computer from ancillary information of the data such as start grid cell location, grid cell size, number of cells in a row etc.

Vector: In vector data model, geographic data are represented as entities. A line is represented by a set of points. By joining these points, finite number of line segments are formed and approximately represent the real line. In comparison a real line is made of infinite number of line segments. Cartesian coordinate system, latitude and longitudes etc. are used to represent the locations in the data.

Comparison of Raster and Vector Models

In raster data models there is information loss when data are changed to a grid cells. All the grid cells within the area are required to be stored. Thus, the raster model requires large storage space in computers. The lines and boundaries are represented in stepped manner. This representation does not show true shape of the boundaries and lines. Advantage of this model is that it is simple to represent in a GIS. In vector data model, the lines and boundaries are represented in their true shape. The data require less storage space in computers. Disadvantage of this model is that it is complex to represent in GIS.

In general, a GIS is based on either a vector or raster data model. Apart from basic data model, few capabilities of other data model is also provided in the commercial GIS. In few Image Processing and GIS software, vector capabilities of other commercial GIS are integrated seamlessly in the system/ module.

4.0 Data Structure

Vector Data Structure

Spaghatti :This type has its origin in drawing software. In this structure, lines are drawn as they are drawn manually. In manual drawing, there is always some under shoot/ over shoot of the lines at there junctions.

Topological: In this structure, there is no over shoot or under shoot at the junction of the lines. The coordinate of the point where the line are supposed to meet or intersect are made by design same.

Raster Data Structure

Raster: This method requires large storage. In this, all grid cell values are stored. The locations of the grid cells are not recorded. The ancillary information is used to know the geographic location of the grid cells.

Run-Length Encoding, Chain Encoding And Quad-Tree: The methods are useful for reducing the storage requirement. They are useful when the maps are simple (having fewer mapping units). When maps have large number of mapping units, these data structures will require space even larger than former structure.

5.0 User Interface

User interface is a method of interaction by a user with an application software. There are many operations/ functions built within the application software. These operations are executed through the interfaces provided. In earlier software, all operations were executed through typing commands at the operating system prompt. With the developments in computer science, the user interfaces are now evolved as a discipline in the computer science. User interface is important to all application software on all operating systems e.g. Unix, Microsoft Windows etc.

Menu: In the menu interface the user works with hierarchical cascading operations lists. At the end of the hierarchy actual function or operation resides. The GIS operations are classified in different groups and a hierarchy is established. Some operation may reside at two or more locations in the menu tree. For example the functions dealing with file operation e.g. opening, creating, copying, renaming, deleting etc. may form one group. The vector operations e.g. creating, deleting, joining, renaming, changing shape, breaking etc. may form another group.

Icon or Graphical User Interface (GUI): 'A picture speaks thousand words.' Icon or GUI uses buttons with pictures or figure pasted on top of them. These figures/ pictures have some logical relationship with the operations. They occupy less space than text. The operations are easily understood.

Command lines: This method is still persisting in the applications. Under this interface, the user writes a command, and presses 'enter' or 'return' key on the keyboard to execute the command. The parameter/ values can also be supplied for the command using key board.

Script/ Model: Script is a method of automating the processing of the geographic data. The script has advantage when the processing of data requires numerous steps or when same processing is repeated for different data sets. This is achieved through use of run time parameters in the script.

Run time parameter are specified during or prior to the execution of the scripts. In the later method parameters are specified in the script run command. The parameters are also supplied to the script interactively. Script can be written in text or graphical form. For scripts special languages are used. The examples of the scripting languages are EML (ERDAS Macro Language), AML (Arc Macro Language), script language of ILWIS software etc.

6.0 Map Projection

In GIS, the maps are geo referenced using a map projection. A map projection is a method to represent the spherical surface of the earth on the two dimensional surface. The surface of the earth is not truly spherical in shape. It is bulging at the equator and flat at the poles. This shape is called geoid.

A hollow cone or cylinder or a plane is placed on the geoid. These objects are called developing surfaces. From a point, rays are drawn towards the geoid, these rays are extended up to the enveloping surface. The enveloping object (cone and cylinder) is later spread in to two dimensional surface. The points at which the rays intersect the developing surfaces represent the points on the geoid. Polyconic, UTM, equal area are few commonly used systems.

Polyconic: The projection system was developed by Ferdinand Hassler in 1820. It is suitable for north-south oriented maps. It uses infinite number of cones as developing surfaces. The parameter needed for the projection, are central meridian, false easting and northing. Central meridian is generally chosen a meridian passing through center of the study area. The parallels are circular arcs and the meridians are radial lines. This projection is used by the Survey of India for topographic mapping at the mapping scales of 1:250000 and 1:50000.

Universal Transverse Mercator (UTM): In UTM, the earth surface is divided in 60 zones of 6° each. The parameters are pre defined for each zone. The whole earth between 84°N to 80°S is covered by the system. It is suitable when the depth of the map is 3/5 the width of the map. False easting is generally 500000 m. False northing for northern hemisphere is 0 and for southern hemisphere is 10000000 m. Accuracy of the projection is 1 in 1000. The projection results in rectangular graticule mesh.

Albers Equal Area Conic: This projection is used for small-scale maps (1:1 m. or smaller). It requires two standard parallels.

7.0 GIS Data Transfer Standard

A data transfer standard are necessary to facilitate the transfer of the geographic data between computer systems. For example ASCII data are acceptable in Windows to the Unix platforms. Few commercial data formats that are widely accepted as data transfer standard are DXF, e00, ERDAS GIS etc.

DXF: This is a data format that writes the vector data in ASCII. This was developed for transferring the engineering drawings (Cassettari 1993).

Arc exchange format (e00): This format is data transfer format used in Arc/ Info. The geographic data are written in ASCII.

ERDAS GIS, LAN and IMG: These are widely used data format for raster data. The data are in binary.

8.0 Data Input

Data input is getting the geographic data in a computer in a format, compatible to the software or which can be made compatible. The data input is done using digitizers, scanners, computer files or manually through keyboard etc.

Manual entry: The vector point data and attribute data are entered manually through key board. For example the location of groundwater wells, their identification, depth etc. can be entered in a spreadsheet program using keyboard.

Digitizer: The Vector point, line and area data can be entered in computer using a digitizer. The data are entered by manually moving the cursor over a map and pressing cursor button.

Scanner: The line, point or area maps can be scanned through scanners. The output formats are TIFF, BMP etc. The output from scanners are always in raster format.

Computer files: GIS data or attribute data may be already available in computer files and these can be used. Example of these data are commercial or other GIS and attribute data, data obtained from processing of remote sensed data etc.

9.0 Import

Import function converts input data in to the internal format of the GIS software. The input data are often in standard data transfer format.

10.0 Georeferencing

Georeferencing is registering a geographic data to another map or geographic coordinates. The digitized data, scanned data are often geo referenced at the time of data input. Other data are geo referenced in the GIS.

11.0 Creating Topological Data Structure

For a topological vector data structure, many errors in vector data are to be removed. The errors are listed here.

Dangle or dead end in segments: This error if any is removed only for the 'area' geographic data, since there should only be closed polygons in this type of geographic data.

Intersection without node: At all intersections of 'line', node should exist in the topological data structure.

Self overlap: This error occurs when a line intersects itself.

Double digitizing: A 'line' can be digitized twice. This causes multiple intersections. This is corrected by deleting one line.

12.0 Digitization Techniques

Reference: Before the digitization is done, the map is placed over the tablet. The map is required to be registered with the tablet. For this purpose, minimum three control points are specified. The ground coordinates are given for the control points for referencing. The digitizer coordinates/ other coordinates of the control points are input in the referencing procedure by placing the cursor at the point and pressing cursor button. There after, the input stream of the digitizer coordinates is converted to the map coordinates automatically in the GIS.

Auto snapping: This is required in topological data structure. A line is automatically connected to another line or node.

Sampling: During digitizing, it is possible to limit the number of points that are accepted by the software from the stream sent by the digitizer. This is done by specifying a tunneling distance or setting limiting time or limiting count or number between samples. In tunneling a width across the line is specified and the intermediate point out of consecutive three points is accepted if it lies outside the tunnel width, otherwise it is rejected.

EDIT

Edit allows changing the existing geographic data. Any vector geographic entity can be deleted, modified and names or attributes can be changed/ modified. Pixel values in a raster GIS also can be changed using edit operation.

13.0 Data Management and Storage

In data management and storage, many operations on the geographic data/ attribute data are done. These are storage, renaming, deleting, copying etc of the data files. The operations can also be done at the operating system prompt after exiting the software interface.

In GIS, many physical files are created for a geographic data. Example of these physical data files are file for actual geographic data, ancillary data, statistics files etc. These are associated files and have no meaning independently. Ideally, a GIS should manage (rename, delete, copy etc.) them as a group i.e. through a single command, name of all associated files should change. This can only be done efficiently through GIS storage and data management functions.

The operation for deleting the geographic data should be done with caution as the data are deleted permanently in most of the commercial GIS when the data management functions are used.

14.0 Data Manipulation

Data manipulations are computations done on the data and editing them. In the data manipulation data statistics e.g. count, average, standard deviation etc. are also computed.

15.0 Measurements and Statistics

Point data: Geo statistical computation namely the semi variance for specified lag distance can be computed from point integer or real values.

Line data: Number and lengths of the lines, rose- diagrams are obtained.

Area/ polygon data: For area data, the area statistics and perimeter are computed.

16.0 Summary

The area statistics of one map for the zones or regions in another map is estimated. The statistics is presented zone wise. For example, given the land use and sub watershed maps for a basin, the land use statistics for the each sub watershed is obtained using the summary.

The mean, median, minimum, maximum etc. are also given for each zone. These parameter are useful in summarizing the value maps. For example, summary of the slope map using the sub watershed map as a zone map will provide mean slope for each sub watershed.

17.0 Analysis

Overlay

In map overlay, two or more maps are combines using certain mathematical/ logical (conditional) functions.

Matrix or Cross

Matrix or cross is a specific overlay operation on two maps. The output map contains values for all combinations of the input map values.

Intersection

Intersection operation is used for overlay of two different types of the geographic data e.g. point and area or line and area. Examples of this operation is finding a sub set of ground water wells data for the wells located inside a basin, extracting village names for the flood affected area, extracting drainage lines in a basin etc.

Identity

In this technique, output shows details of both the maps within only the extent of one map.

Union

In union, output shows details of both the maps.

Search or Buffer

In this operation, a buffer is created around the geographic data e.g. point, line or area. The extent for the buffer can be specified. For examples using the operation a map showing the buffer zones around lake and rivers, roads, rail etc. can be created.

Neighborhood

Neighborhood is a function that operates on neighboring grid cell. The maximum, minimum, predominant or majority, density, rank etc. can be computed on the neighbor cells. The operation can also be done on any set of the neighbor cells e.g. all neighbors including the central pixel, 8-neighbors and 4-neighbors or any user defined sets. For example, the majority is applied to remove noise pixels in maps prepared through digital classification of remotely sensed data.

Morphology

Morphological operations are specific neighborhood operations on binary images. The operations described here and other few operations are used in applications e.g. thining, thickening, boundary detection etc.

Dilation: In this operation size of the binary image is increased by specified number of pixels on all sides including the islands. For this a maximum neighborhood operation is applied.

Erosion: In this operation the size of the image is reduced by specified number of pixels on all sides including the islands. For this a minimum neighborhood operation is applied.

Opening: Two operations namely erosion and dilation are applied in succession for opening function. This removes thin connecting corridors between mapping units.

Closing: Two operations namely dilation and erosion are applied in succession for closing function. This connects mapping units separated by thin gaps.

Distance

Distance function gives distance from point, line or area. The diagonal distances are nearly 1.414 times larger than horizontal and vertical distances in a raster distance operation.

Classification

Classification is a method of information retrieval from the value type data such as soil pH, slope, elevation maps, rainfall isohyetes etc. The upper and the lower bounds of values are specified for the groups. Any value falling within these bounds is assigned to a group.

Reclassification or Recoding

In reclassification each class is assigned a new value or class. In this two or more classes can be assigned one class or value. For example from a soil class map, soil-pH map can be prepared, from a 2nd order land use map, 1st order map can be prepared etc.

Dissolve

This function is needed in map reclassification in a vector data models. For example when a level-1 land use map is created from a level-2 map, the boundaries of the level-2 classes are removed by this operation.

Query

The query can be done on a map or the attribute database. The attributes of a geographic entity can be read by highlighting it using the cursor. On the other hand, a SQL query can be made on the database and the corresponding geographic data can be displayed on the monitor or saved on the disk. By selecting any map object or roaming the cursor over the map, the attributes of geographic object are automatically displayed.

In a river network, the attribute of any river can be read by selected that river from the map. The SQL query is run in data base to select records. The corresponding geographic objects are automatically highlighted on the map. For example, by running query on field containing name of rivers, the color of the queried river changes in the map or if the queried river is not displayed, the display is reset to display the river.

Output

The GIS layers and databases can be output on printers. The output can also be displayed on the monitor screen. The advantages of GIS output are automatic annotations, ease in map design, small response time, saving of the design on computer, automatic updating of the color scheme, printing of map series etc.

A layout of the output is prepared as a first step. In the layout, the location of cartographic elements e.g. map, legend, scale, north arrow, title are specified. The elements are selected for display from available designs.

The GIS layers, their colors and symbols are also designed. The choice of colors can be natural colors e.g. water can be shown in blue color. The symbols are only available in vector data models and not for the raster models.

The information such as pie charts, histograms, proportional circles etc. can also be displayed at specified location. For example, the monthly normal rainfall at rain gauge stations can be displayed as histogram. The resulting layout is saved in the disk or printed. The display color can be modified for maps to be printed.

The maps are dynamically read from the disk. If any changes are made in the component maps after the layout is saved, the changes are automatically displayed in the layout when it is opened later.

The map series tool is useful for printing a large area map in sections. The extent of the map and the size of the sections are input to the map series tool. Using these input, the layout for individual sections are created from the master layout.

18.0 Visualization

Geographic data are visualized using video display unit (VDU). The data may be visualized as map or three dimensional views. Both of these visualizations are described here.

Two Dimensional Visualization

In map visualization the geographic data can be displayed on the VDU. The data may be displayed in color, gray tone, as shades or as symbols. The area or raster data can be displayed in color or gray tones. Different classes or values can be assigned different colors or gray shades. For value maps e.g. elevation, soil pH, groundwater depths etc., the color may be assigned automatically. In these standard color schemes the colors are assigned from blue to red and the intensity in each color changes from low to high as the values in the geographic data increases. In another scheme, the colors are assigned among violet, indigo, blue, green, yellow, orange and red. The intensity in each color is increased with increase in the geographic data values.

In the area maps (vector), color, gray tones, shades and hashing are used. The hashing and shades can be assigned different colors. In points maps, points are displayed as symbols. The size and color can be assigned to the symbols. For line maps, lines may be given different types, thickness and colors e.g. a line may be dash and dot type, double lines alternately filled with black and white color, single line with two small vertical line segment pairs placed at regular spacing etc.

Multiple maps with same geo reference may be displayed one over another. Any map can be put on top of other map/s in the display. Maps can be zoomed in or out and scrolled. If the maps displayed in different windows have same geo reference, the displays can be linked. In the linked displays, scrolling a map, scrolls maps displayed in other windows automatically.

19.0 Digital Elevation Model (DEM) or Digital Terrain Model (DTM)

DEM or DTM deals with the topographic elevation of the ground surface. Conventionally, the ground elevations are available in the topographic maps as contours, spot heights etc. In GIS, the information is available similar to this conventional form as vectors or points. Apart from this the data are also represented as raster and TIN model. Since, water flows under gravity from higher elevation to lower elevations, the DEM are very important input in hydrological studies. Using DEM, the surface is routed from upstream points to the catchment outlet. Other physical variables e.g. slope, aspect, drainage network, subwatersheds are also derived from DEM automatically.

The DEM for areas in USA are prepared by USGS as profile files (West-east) at 30 m X 30 M resolutions and 7.5' X 7.5' blocks. Another DEM for USA by USGS is at 3" X 3" resolution and in 1° X 1° blocks. ETOPO5 is a global gridded elevation and bathymetry at 5' X 5' latitude and longitude prepared by NOAA. HYDRO1k is a hydrologically connected DEM.

The raster and TIN forms are useful input to hydrological models. Thus, the conventional forms of DEM such as vector, point are required to be converted to these forms. This is done through the interpolation techniques in GIS.

Triangulated Irregular Network (TIN): In this a network of triangles is created by joining point with techniques known as Thiessen polygons. At each apex of the triangles the height is known. The ground is represented as a plane surface with in a triangle. Input data to the model are spot height and contour data of the topographic maps. The triangles are so drawn such that sides of the triangles are the lines of slope or aspect change and the apex are the points where spot height is known.

20.0 Interpolation

Point:

Nearest neighbor: This technique gives horizontal surfaces around each point. This technique is generally used for interpolation of the rainfall data. This provides weight to be used for the rain gauge stations, for finding average rainfall over a watershed.

Weighted average: A weighted average is taken of point elevation falling within a specified radius from the interpolation point. Weight is an inverse function of a distance between the point of unknown value (interpolation point) and that of known value (data points).

Weighted surface: Weighted surface fits a surface of n-degree using the elevations at points within a specified radius from the point of interpolation.

Trend surface: Trend surface fits an overall surface with the input data. It provides the global direction in which the variable changes.

Kriging/ Geostatistical method: Kriging is a mathematical technique developed by D.G. Krige for application in mining industry in South Africa. The experimental semivariogram is calculated from the data. A standard semivariogram equation is fitted to the experimental data using regression technique. This fitted curve is used in the interpolation. For groundwater data the spherical theoretical semivariogram model is best suited.

21.0 Terrain Visualization

Many techniques are used to represent the terrain on VDU. The techniques use simple representations such as contours to sophisticated representations in which observer feels as if he is part of the scene such as virtual reality.

Contouring and hypsometric representation: The techniques have been in use in topographic maps, physical maps etc. A contour is a line of equal elevation. In hypsometric representation, the areas between contours are shaded with different colors. In GIS contour data are represented by vector model and the hypsometric data are represented by vector or raster models.

Hillshading: The illumination on objects is calculated. A light source is assumed over an object. The tone or color of the object is calculated using DEM. This type of shading has been in use with conventional paper maps and is also available in many GIS.

Perspective view: A perspective view involves calculation of hidden points. The technique involves specifying location of the observer, look direction, vertical exaggeration, angle of view or panorama etc. In The flat areas the height is exaggerated for a better perspective view.

A perspective view is created using wire net or profiling techniques. The adjacent profiles of the terrain are joined together to produce the perspective view. Over the perspective view, optionally, the GIS layers can be draped.

Fly through: In this technique, many perspective views are generated with several points as observation points. These views are animated to create an impression of the flying over the terrain. The direction and the speed of the flight, height etc. can be defined interactively.

True three- dimensional view: In a true three- dimensional view, by changing the height or the distance between viewer and the objects, the details visible are also changed.

Perspective cartography and texture mapping: In the perspective cartography and texture mapping, an image or symbols for important buildings, bridges, built up areas are overlaid on the perspective view.

22.0 Hydrological Modelling

SCS Curve Number Technique

The direct runoff volume is computed from agricultural/ urban watersheds etc. using USGS SCS runoff curve number method. In applying these methods, the factors are averaged over the watersheds and the empirical equations are used to estimate the hydrological quantities. In other approach, the estimations are done at each raster grid cell and the result is averaged over the watershed to estimate the hydrological response of the watershed. The method is combined with the unit hydrograph method, to obtain direct runoff hydrograph (DRH) from catchments. The input required are catchment physical characteristics, design storm and the curve number for the catchment. The curve numbers given soil cover complexes for rural as well urban areas. The method has been applied for level-1 land use maps derived from remotely sensed data.

Suitability/ Vulnerability Study or Predictive Modeling

In these studies the weight factor maps are overlaid and the resultant map gives information regarding suitability or vulnerability of an area for an application. In other approach a composite map is prepared. In the composite map, the units are assigned the suitability or vulnerability values or indices. Example of this technique is Sediment Yield Index (SYI) method developed by All India Soil and Land Use Survey Organization, Department of Agriculture for prioritization of watersheds with objective of taking steps to control soil erosion from catchments.