TRAINING COURSE
ON

APPLICATIONS OF REMOTE SENSING AND GIS IN WATER RESOURCES MANAGEMENT

ORGANISED UNDER
HYDROLOGY PROJECT PHASE II

AT
NATIONAL INSTITUTE OF HYDROLOGY, ROORKEE

MARCH 08-12, 2010

LECTURE NOTE

FUNDAMENTALS OF GIS AND APPLICATION IN HYDROLOGYCTION AND PRINCIPLES OF REMOTE

By

D.S. RATHORE
NIH, ROORKEE

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

FUNDAMENTALS OF GIS AND ITS APPLICATION IN HYDROLOGY

1.0 INTRODUCTION

GIS is a computer-based system that is used in input, output, storage, manipulation, retrieval and analysis of spatial data. These systems consist of computer hardware and software. GIS are increasingly being used in applications in natural resources, tourism, transportation, trade and commerce etc. GIS is also integrated with modern technology of remote sensing and GPS. GIS are now integral part of hydrological models.

History

GIS has basis in manual overlay operations done as early as in 1912 to 1969. In 1958, computer based cartography initiated in University of Washington, which culminated in development of first general purpose mapping software in 1960s. Canada GIS (CGIS) is also cited as first GIS and was developed. Other GIS developed were SYMAP (1964), SYMVU (1960s), GRID (1960s), ODYSSEY (1970s) etc. In 1960s, US Bureau of the Census created census information of USA on GIS. In 1970s, Environmental System Research Institute (ESRI) was founded. ESRI combined ARC and a relational database management system (RDBMS) INFO to develop ARC INFO GIS package in 1980s. During these decades, software e.g., MAP, Intergraph's CADD and mapping software, MOSS BLM, GRASS (Army Corps of Engineers, National Park Service), SPAN (Tydac Inc.) etc. were developed.

Present day commercial/ open source GIS are Arc GIS, GRASS, MapInfo, ERDAS, IDRISI, ILWIS etc. GRASS (Geographic Resource Analysis Support System) is high-end open source software. Arc GIS is modular high-end commercial software. ERDAS, ILWIS and IDRISI have image processing and GIS capabilities.

Advantages and limitations

Using GIS, it is possible to overlay large number of maps. Conventionally, manually overlays are prepared. This process is cumbersome and error prone. In GIS retrieval of information is faster as it is done through computer. Conventionally, maps are browsed to retrieve information. In GIS information retrieval is much easier and is done automatically. In GIS, interactive/ virtual output may be prepared. The virtual output is automatically updated, if the component maps are revised. In convention method, hardcopy output is prepared. Updating of such maps is difficult. The map is required to be redrawn. Annotation is clumsy in hardcopy maps. Thus, while retrieving information ambiguity may arise. Also all features may not be annotated in paper maps. Thus, attribute information e.g. names etc. for some of the features are lost in paper products. In GIS, information is stored in tables and is linked to geographic features and thus is not limited by availability of annotation space/ color/ symbol etc. If multiple maps are prepared for same area e.g. watershed, land use, geomorphology, common boundaries are drawn manually and may not match in different maps. In GIS common boundaries are once digitized and are available to all layers. Once GIS map layers are prepared, any number of maps can be designed. Handling of paper maps is difficult.

Data capture or input is costly in GIS. Commercially available paper maps may be cheaper than GIS layer. This is because of high cost of data capture in GIS. Use of GIS requires investment in computers, software and training. GIS handling requires trained manpower. In GIS data are required to be converted in to native format of GIS software. In suitable import/ export functionality is not available or the format is obscure/ unknown, data may be unusable. GIS software should have proper functionality as desired in an application. For example, in transportation applications, network analysis function should be available. For hydrological modeling, DEM analysis functions should be available.

1.1 SPATIAL DATA

In GIS, maps are called spatial data. Information on paper maps can be input is GIS as spatial data. Example of spatial data are stream network, well locations, villages, cities, topographic contours, spot elevations, roads, land use, soil, geology, hydrological investigation locations, hydrological response units etc. Spatial data are classified in to three types namely area, line and point. Areas are spatial data that are represented as closed figures e.g. forests, lakes, Thiessen polygons etc. Lines are spatial data that are represented as lines e.g. forest boundaries, lake boundaries, contours, stream network, roads etc. Point spatial data are represented as points on maps e.g. well locations, rain gauge stations, spot elevation, villages etc. The data is scale dependent in some instances e.g. on small-scale maps a city will be represented as point data, where as on large-scale map, it will be represented as area data. A closed line data can be converted easily to area data in GIS. For example forest boundary data can be converted to forest land-use map. Point data cannot be converted in to area data. But reverse is true i.e. area data can be converted as point data. For example, if city point data is captured, city polygons cannot be obtained from this point data. City polygon data will be required to be captured separately.

1.1.1 Spatial Data Representations

Spatial data are represented in two mainly ways in GIS namely raster and vector. These data representations can be translated in to one another, albeit with some information loss. In raster, spatial data are structured as grid of cells or pixels. Their row and column numbers addresses the cells. In many distributed hydrological models, processing is spatial data are represented in this form and hydrological computations are also done in this form. This is a native representation for remotely sensed data. Satellite data are captured/ resampled as pixel (picture elements) or grid cell. Thematic maps are prepared from these data through digital image processing. These maps are available in raster representation. In vector model, spatial data are represented as coordinate points. For example, point data is represented as a pair of coordinates. A straight-line data is represented as two pair of coordinates, representing end points of the straight line. A curved line is represented as finite line segments. Area data are represented as line data with some additional information e.g. centroid, adjacent areas etc.

Comparison

In raster data, points and lines are represented with finite area and finite width and thus is not a natural representation. Lines has jagged or stepped appearance. In vector model, points and lines have infinitesimal area and width respectively. Lines are smooth curves. Raster data require large storage space. Vector data require small storage space. Thematic maps prepared from remotely sensed data are available is raster form and are often processed as such. Many hydrological models use both the representation. For example, thematic maps of catchment variables and hydrometeorolgical measurements are prepared in raster form. Stream network is processed in vector form etc. In raster form, value of many catchment variables is scale dependent. For example, average slope of catchment reduces with increase in raster grid size. Thus, results of un calibrated hydrological models will differ at different raster grid size used in parameter derivation. In most of GIS, the representations coexist. For example, it is better to capture spatial data from conventional thematic maps, through visual interpretation of remotely sensed data etc. in vector form. Thematic maps from digital processing of satellite data may be obtained in raster form. Data can be transformed in to one another as and when desired.

Topology

Method of representing vector data is called its topology. A line consists of two nodes and one or more vertices. Nodes are end points of the line. Lines also have directions. Thus, nodes are referred as 'from node' and 'to node' depending on direction of the line. Areas are represented by 'left area' and 'right area' of each line.

1.2 DATA FORMAT

There are many formats prevalent for images, raster and vector GSI data. Image file formats are normally used by non GIS applications. For use in GIS, data in these formats are required to be imported. The image file formats may or may not have projected coordinate system.

Image file formats

GIF (Graphics Interchange Format): This is copyrighted format by 'CompuServe' for image data. A data compression scheme similar to lossless LZW compression is used. Georeference information of raster data is not stored.

TIFF (Tagged Image File Format): The format is designed by Adlus and Microsoft. For types of images are supported namely bitlevelm gray scale, color- mapped (256 color) and RGB- based (three layer or band for full color). The compression type can be uncompressed, RLE or modified Huffman. A format for storing georeferenced images is also designed and is called 'geo- tiff'.

JPEG (Joint Photographic Expert Group): The format uses lossy compression technique. Compression of 20:1 can be achieved. The format is designed for grey- scale and color images. Georeference information of raster data is not stored.

PNG (Portable Network Graphics): This format uses lossless compression and the compression technique used is not a patented technique. This format is suitable for images with line drawings rather than true color images.

GIS file formats

NTF (National Transfer Format): This is developed in United Kingdom 9UK) for transfer of digital cartographic data in 1987 and updated in 1989. It is format for both raster and vector data. Levels are defined from 0 to 5 for raster, simple vector, complex vector, topology and user defined formats.

SDTS (Spatial Data Transfer Specifications): The data exchange format standard is developed by US Digital Cartographic Data Standards Task Force (DCDSTF) in 1988 for both raster and vector data.

e00 (Arc Exchange Format): This is exchange format developed by ESRI for vector data.

Arc coverage: This is a format for vector data developed by ESRI. The format is also used in ERDAS imagine.

Shape: This is a format for vector data developed by ESRI. The format is also used in ERDAS imagine. A shape file is a collection of files with extension shp, shx, dbf, prj etc. File with shp extension stores actural geographc data. File with shx extension stores index. The file with dbf extension is a dbase III file which stores attribute data. The prj file stores projection information and is a ASCII file. Shape files do not store topological information.

Img: This is a format for raster data developed by ERDAS Incorporated (now Lyca).

DXF (Drawing Exchange Format): The format was designed for data exchange between Autocad and other Cad packages. This is a de facto standard for exchange of vector data.

TIGER (Topographically Integrated Geocoding and Referencing) and GBF- DIME: US Bureau of Census developed these formats in respectively 1990 and 1980. Former is a topologically structured format.

DLG (Digital Line Graph): US Geological Survey (USGS) developed the format for transfer of fully structured data from National Cartographic Database. DLG-3 and DLG-E are types of this format.

BSQ/ BIL and BIP: These are generic data formats and are mainly used for remotely sensed data and some times for GIS data e.g. in ILWIS software. In BSQ (Band Sequential) format, each band of remotely sensed data is stored in separate files. In BIL (Band Interleaved by Line) format each band of a scan line are stored in alternate records in single file. In BIP (Band Interleaved by Pixel) group of pixels (e.g. two pixels) are stored for each band, are stored alternately.

1.3 DIGITAL ELEVATION MODEL (DEM)

Topographic elevation data in GIS are called DEM. These are represented in GIS in various manners namely contours, raster, TIN (Triangulated Irregular Network). Contours are conventional representations of DEM and are used in topographic maps. Contours are equal elevation lines. Normally, equal interval contours are drawn in topographic maps to represent topography. For example in 1: 250,000, 1:50,000 and 1: 25,000 scale Survey of India (SOI) maps, contours are at 100, 20 and 10 m elevation interval. Ridges, valleys can be interpreted from these maps. DEM in Raster and TIN representations can be used in deriving topographic information such as slope, aspect and can also be used in hydrological calculations e.g. stream network delineation, topographic index, flow routing, up steam contributing area etc. and in turn in hydrological modeling.

1.3.1 TIN

In TIN model, elevations at the vertices of triangles are used to compute elevation at interior points of the triangles. Using elevation of the vertices of a triangle, a planner or higher order surface can be fitted. The surface can be used to derive elevation at points inside the triangle. TIN model requires Delaunay triangulation. In this, constituent triangles are as equilateral as possible. Circumcircles of the triangles include no other point of the triangulation. Triangulation is performed first by constructing Voronoi diagram (Thiessen polygons). Points included in adjacent polygons are joined to create Delaunay triangulation. Voronoi diagram is drawn using proximity analysis.

1.3.2 Interpolation

Interpolation is a technique of determining unknown value of a variable at location from known values at other locations. Interpolation can be used for any spatial variable e.g. topographic elevation, pH, SAR, pollutant concentration, groundwater depth and level, population etc. Known values can be at point, line of area locations. Point data can be spot heights, pH, pollutant concentration etc. Line data can be topographic contours etc. Area data can be population density in regions etc.

<u>Voronoi diagram or Thiessen polygons or nearest neighbor</u>

To determine basin wide average rainfall, this method of interpolation is widely used. The diagram is prepared by proximity analysis. The Thiessen polygon map is intersected with the catchment map. Area of a Thiessen polygon corresponding to a raingauge station in this intersected map is used as weight in finding weighted average rainfall for the catchment.

Distance weighted averaging

In distance weighted averaging, a weight of inverse of distance function is used. Distance function is nth power of distance. Thus, a higher weight is assigned to values closer to the interpolation location. At any point values are estimated as weighted sum of known values at selected locations. The selected point can be as follows:

- All points within a given range
- Specified number of closest points
- Specified number of closest points within quadrants/ Octants etc.

Surface fitting

A n- degree polynomial surface can be fit between selected known values. The points on this surface represent interpolate values. The points to be used for interpolation can be selected in similar way as that in distance weighted averaging method.

Kriging

Kriging is a statistical technique called best linear unbiased estimator (BLUE). Spatial variables have three components namely drift or structure, small variations and random noise. First component depicts general trend of the data. Second component represents small variations from the general trend. These variations are random but spatially autocorrelated. Third component depict random values that are not spatially autocorrelated. Kriging technique is best suited for interpolation of pollutant concentration, geological and mining variables e.g. grade of ores etc. In these data, single smooth mathematical equations are not suitable. The technique is based on assumption that values in neighbourhood have generally higher correlated. For example elevation in plain area is generally lower than that for hills and varies less abruptly. Apart from the estimate of values, error estimates are also provided in kriging technique. In presence of large random noise in data, good semivariogram is not obtained and this results in deterioration in interpolation quality.

Semivariogram

Semivariogram is a plot of semivariances and distances of the samples for which the semivariance corresponds. Semivariogram is also referred as variogram. When variogram for all separation distances are plotted, the resulted variogram plot is called raw variogram. For 'n' data points, the resulted points in the raw variogram will be 'n*(n-1)/2'. Raw variogram show cloud of points. A representative variogram or experimental variogram is more useful for fitting theoretical models and doing kriging. In computation of this, the separation distances are grouped. Plotting positions for these groups are mean, median, or middle of the class intervals. Distance intervals for these groups at smaller distances can be smaller.

For mathematical formulation of kriging interpolation, a theoretical models is fit for experimental variogram. Models e.g. Gaussian, exponential, power etc. are used. Gaussian model has parabolic nature at origin. Exponential and power models are linear at the origin. Parameters of the model are sill, range and nugget. The sill is equal to the variance of the data. Due to experimental error and micro level variations, experimental variogram may not pass through origin. This property is modeled by Nugget. It is modeled through a discontinuity at the origin.

Validation

'Jack- knifing' can validate the model. A subset of data is used in kriging. The complementary datasets of these subsets are used in validation of selected model. Residuals and standard errors are estimated at all discrete locations. The mean and variance of these residuals should be theoretically 0 and 1 respectively. Normally acceptable deviations of these measures from their true values may be utmost 0.15 to 0.20.

Linear contour interpolation

In linear interpolation from contours, distance map is estimated from contours. Based on distances at an interpolation point towards two nearest points, value is interpolated linearly. The value at the contours is retained in the final map. Distance function can be used in estimating distances.

Ray method

In this method rays are drawn in four, eight or sixteen directions from interpolation points. A value of the contours at points where the rays intersects the contours is determined. Using two nearest values, interpolation is done. Average of all interpolated values provide required interpolation.

1.4 GEOREFERENCING

Any location on the earth surface can be assigned geographic or Cartesian coordinates. Geographic coordinates are expressed in latitudes and longitudes, called graticles. Cartesian coordinates are expressed as abscissa and ordinates (X and Y coordinate respectively). Cartesian coordinates can be a plannar coordinate system or a projected coordinate system. In projected coordinate system, earth spheroid is projected to a developable surface. Earth surface has spheroidal shape. This shape is approximated as a ellipsoid in map projection. Along with a datum, the map projection system is defined for the earth surface or part of it. A map projection system can be a standard system in which one developable surface is used or it can be special system. Projected coordinates allow measurement of line and area and thus are preferred systems.

In georeferencing, through transformations, the coordinates are assigned to images rows or column values. A table of row and column values and corresponding geographic coordinate values are prepared. An inverse transformation matrix is computed for transformation of geographic coordinates to rows and columns. The pixel size of the georeferenced image is selected. For the pixel size, a grid of georeferenced image is prepared. The points on this grid are transformed in to rows and columns locations. These transformed grid location lie in the neighbourhood of input row- column grid points. A interpolation technique, e.g. nearest neighbor, bilinear or cubic convolution etc., is used to find values at transformed grid locations. In these techniques one, four and 16 nearest values are used respectively.

Following is an example of transformation polynomial:

$$x' = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6xy^2 + a_7x^2y + a_8x^3 + a_9y^3$$
$$y' = b_0 + b_1x + b_2y + b_3xy + b_4x^2 + b_5y^2 + b_6xy^2 + b_7x^2y + b_8x^3 + b_9y^3$$

In bilinear interpolation, first linear interpolation is done in X direction, at X location of the transformed grid. Four neighbours will have two X direction resulting in two values. These two values will be values at two Y locations. These values are interpolated linearly for transformed grid Y location.

In cubic convolution, 16 neighbours are used in interpolation with interpolation equation used as follows:

$$p(x,y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j$$

1.4.1 Map Projection

Map projection is transfer of positions on earth to corresponding points on a flat sheet of paper. Because of the near spherical shape of the earth, this transformation involves approximations and is not distortion free. Distortions occur in lengths, angles, shapes and areas. Scale is a ratio of length on map to its counterpart on the earth. Since large size of features on earth surface, a scale is needed to draw these features on a small sheet of paper. Earth shape is assumed spherical or spheroidal. An intermediate plotting surfaces namely cylinder, cone or plane is used in projections. Corresponding projections are called cylindrical, conical and azimuthal respectively. Distortions occur in projecting earth surface on to intermediate plotting surface. Ideally, areas distances, directions, angles and shapes should be preserved. In reality, few of these properties are preserved. Based on application, choice is made as to which propertied are to be preserved and appropriate map projection is selected. In areal distortions area of a figure may increase or reduce. In linear distortion length and its curvature may change. In angular distortion an angle may increase or decrease. In shape distortion, a square may become parallelogram, rectangle or may have curved boundaries or both. A point may be distorted in to a line. In equal area projections, area of a figure is preserved. In the process distortions are introduced in distances and angles or shape of figures. In conformal projection, shape is preserved. In this process, areas figures are distorted. Projections with these contrasting properties are called equal area and conformal projections respectively. With different orientation of intermediate plotting surface, it is possible to obtain different projections. Azimuthal projections are called polar, equatorial or oblique depending on point of contact of plane falling on poles, equator or at intermediate latitude. For conic and cylindrical projection based on orientation of axis of the plotting surfaces, the projections are classified. When the axis coincides with earth's polar axis, perpendicular to it and lying in equatorial plane and is oblique to it, the projections are called regular or equatorial, transverse and oblique respectively. The plotting surfaces can also be tangential or secant to the earth surface. Normally for projection, mathematical approach is used. Various map projections are listed in Table 1.3.

Table 1.3 Map projections

Projection	Type (Plotting surface)
Equal area	
Albers, one standard parallel	Conic, tangent
Albers, two standard parallels	Conic, secant
Lambert zenithal	Azimuthal
Cylindrical equal area	Cylindrical
Mollweide	Special
Interrupted	Special
Conformal	
Lambert conformal, one standard parallel	Conic, tangent
Lambert conformal, two standard parallel	Conic, secant
Stereographic	Azimuthal
Mercator (e.g. Universal Transverse Mercator or UTM)	Cylindrical
State plane (US)- Transverse Mercator- secant and Lambert conformal with two standard parallels	-
Military grid (US)- UTM and universal polar stereographic (UPS)	-

1.5 GIS OPERATIONS

1.5.1 Input

Digitization

Digitization is done on- screen to create/ edit GIS objects in vector format.

Digitization errors for area objects

Following errors occur while digitization of area objects:

Dangle: Additional lines are some times digitized that are not areas. These lines appear as dangles.

Overshoot: Areas are closed figures. Sometimes, line is extended beyond closed figure. These are called overshoots.

Undershoot: Sometimes, line stop before closing of the area. These are called undershoots.

Self-overlap: A line crossing itself is called self-overlap.

Intersection: Lines cross each other without a node at the intersection.

Other errors are missing lines, incorrect labels etc.

Data import

Input data are, some times, available in GIS, image formats. These data are converted to native format of GIS through import utility.

1.5.2 Storage

Geographic data are stored in GIS is native format of GIS. For one data, many computer files are created. These files are copied, renamed, deleted wihin GIS. These operations can also be done outside GIS. Attribute data are stored in DBMS. Attribute data are managed within GIS or through DBMS software. Attribute data are linked to geographic objects. External databases can also be connected to geographic data. Data types of geographic objects and their attributes are bit, byte, integer, real, double, text etc.

1.5.3 Analysis

Data analysis involves operations with geographic data and their attributes to obtain derived information, generate query, statistics etc.

Statistics: Statistics e.g. count, length, area, perimeter, shape, centroid, rose diagram etc. of geographic objects can be derived in GIS. For continuous surfaces, average, standard deviation, maximum, minimum etc. are derived. Summary operation produces zonal statistics for a map. For example, land use statistics for watersheds in a basin can be generated.

Mathematical operations: Mathematical operations e.g. addition, subtraction, multiplication, division, exponential, logarithm, absolute, truncation, round off, negative, trigonometric operations can be performed in GIS. For example various component maps in USLE namely R, K, L, S, C and P can be prepared and multiplied using multiplication operation. This operation will multiply these factors for all cells and provide long-term average annual soil erosion for each cell.

Logical operations: Logical operations namely or, and, not, xor can be performed on maps. For example, landuse= agriculture and pH>=8 will result in salt affected agriculture area.

Conditional: If- then-else conditional operational can be performed on maps. For example, 'if 50 < return period <=100 and land use= residential, then vulnerability= high else vulnerability=low' condition gives flood vulnerability map.

Overlay: In this operation, all combinations of classes in two maps are obtained in the resulted map. For example, overlay of soil hydrological soil group and land use/ cover map will provide soil- cover complex map.

Reclassification: Information of geographic object is changed in reclassification. For example soil series map may be changed to soil pH map.

Classification: Classification converts values in to interval. A continuous surface is input and area map is output for the operation. In the output area map, isolines, i.e., line of equal values, enclose the area. Examples of various isolines are contours, isobath, isohyete, isotherm, isobar etc., which represent topographic elevation, groundwater table, rainfall, temperature, pressure etc.

Distance: Distance from a geographic object is estimated. Diagonal distances are nearly 1.4 times that of horizontal and vertical distances.
Search/ buffer: The operation is similar to distance, except that at a specified distance an area geographic object is created.

Neighborhood: Information in eight neighbor, their locations and statistics e.g. mean, mode, median, predominant, minimum, maximum, standard deviation, coefficient of variation etc. are extracted.

Aggregate: Cell size of raster maps can be changed in fractions of half, one fourth etc. using functions e.g. mean, predominant, minimum and maximum.

Query: Query is done by attributes or geometry. In query by attribute, a logical expression is written in attributes and result is obtained. For example land use=agriculture will select/ display agriculture areas. In query by geometry, objects are selected on screen to view their attributes.

1.5.4 Output

The output maps may contain various cartographic elements namely title, legend, graticules or grids, north arrow, scale, annotations, notes etc. In one output more than one GIS layer may be included apart from cartographic elements. When design is saved, it only contains only reference to the layers. Thus, if a layer is modified and designed output map is opened at a later time, the changes are reflected in to the output.

1.6 GIS SOFTWARE

There are many GIS software available both in public domain (free of cost) and as commercial software. Examples of public domain/ open source software are GRASS, OSSIM, QGIS, ILWIS (since July 1, 2007) etc. Commercial software are Arc Info, Map Info, ERDAS Imagine, Geomatica etc. Among these, Arc Info is high end software. Several software e.g. R2V, Surfer etc. have specific GIS functionalities. R2V is useful software for data input from scanned images. Scanned images are converted to vector format using the software. Surfer has many GIS function e.g. interpolation, contouring, output, three-dimensional visualization etc. It is widely used as a spatial data interface for Modflow groundwater modeling software. Some of the main stream GIS software are described below:

ILWIS (Integrated Land and Water Information System)

The software has both image processing and GIS capabilities. It has scripting language. Application/ interface development environment is not available. It supports raster format and limited support for vector format is available. Some of the functions available are topological digitization, polygonizsation, raster to vector conversion (polygon), projection transformation (both water and vector), hydrological processing of DEM etc. The software will be available as free software since July 1, 2007 under GPL. It is no longer supported by ITC, The Netherlands from this year.

GRASS (Geographic Resources Analysis Support System)

It is a modular system and was developed by USA- CERL (US Army Construction Engineering Research Laboratory). Since 1999, it is available under GNU GPL as pen source software. It has many of the capabilities as described above for ILWIS. Interfaces for hydrological models have been developed for the software. It has interface with QGIS software for data input. It is developed under cygwin/ X- window environment under MS Windows.

OSSIM (Open Source Software Image Map)

It has image processing, GIS, photogrammetric capabilities. It was funded by several US government agencies. The program was developed in C++ with GUI and command line utilities.

QGIS (Quantum GIS)

GIS is an open source GIS and support several vector, raster and database formats e.g. shape, geotiff etc. The GIS has capabilities to visualize, manage, edit, analyse data, and compose printable maps. Current version is Enceladus 1.4.0. For editing topological editing for line and polygon data is available.

Arc GIS

Arc GIS software is developed by ESRI and is modular, high end commercial software. The software has several versions namely Arc Reader (free), Arc View, Arc Editor and Arc Info. For each version several extensions are available seperately. Applications/ interfaces can be developed using the software objects provided. Various programming languages e.g. Visual Basin, .Net platform etc. may be used in developing these applications. For hydrological software, several extensions are developed by the third parties and are available from them. It simplifies creation of data layers for hydrological software.

Map Info

Map Info is a vector based GIS and is suitable for projects where attributes data are required to be manipulated.

1.7 DATA BASE MANAGEMENT SYSTEM (DBMS)

Data bases are collection of data put in an organised manner, specifically put in a tabular form. An example of database is watershed database, which contains watershed name, code, area, perimeter, areas under different land use, average rainfall, runoff yield at different dependability, groundwater potential, population, water demand for domestic, cattle, industrial purposes, average slope, stream length, stream slope etc. The information may be collected or derived from various sources e.g. topographic maps, census, remotely sensed, hydrometeorological data etc. Information may be primary or secondary.

For storing, editing, manipulation, retrieval and output of data bases special software are available. These software are called 'data base management system' or DBMS. Examples of DBMS software are MS Access, MS SQL, MySQL, Oracle etc.

In DBMS, information for any object, e.g. a watershed, is stored in a single row or record. Each column contains single information or attribute. For example a column named 'watershed_area' will contain watershed area information only. It is not possible to store along with area, specific notes e.g. source from which area was obtained. For this purpose, another column may be created in the data base where notes may be stored. Also, it is not possible to record multiple information e.g. watershed area from different sources, in a single column. This requires separate columns, e.g. 'W_area_topo', W_area_report" etc.

GIS requires storing of attributes of spatial objects. DBMS are required to handle these data. DBMS functionality can be coded within GIS software or external commercial DBMS may be utilized by GIS. For example, ILWIS software itself handles its attribute data. Arc Info originally utilized 'Info' commercial DBMS. Now, a commercial DBMS of user's choice may be utilized in Arc Info.

Table or relation

A data base may contain several tables. For a watershed, there may be several tables, e.g. for watershed characteristics, dams salient features, conservation agencies etc.

Primary key

A primary key is one or more columns in a table used to identify a row. Primary key has unique values. For example, watershed code attribute may be used as primary key in watershed characteristics table. Since this attribute has unique value, any row can be identified using this attribute. Watershed names can also be used as 'primary key' provided they are unique. Many times, primary key may be generated automatically by the DBMS software.

Relations

A data base is organised in to several tables. These tables have relation between them. The relation may be of several types, namely one- to- one, one- to- many or many- to- many.
One- to- one: Watersheds are delineated at the dam location. In such case, watershed characteristics table and dam information table will have one- to- one relation.

One- to- many: In a basin, there may be several dams. In such case, basin characteristics and dam salient feature tables will have one- to- many relationship i.e. for one basin there can be several dams.

Many- to- many: In a command area and watershed data base, there can be many- to- many relationship i.e. a command may spread across several watershed or a watershed may be under several commands.

Join

Join is an operation used to extract information across several tables. For examples, to get information of dams located within several basins, tables of basin and dam will be joined. Join operation depends on type of relationship.

One- to- one: In this type of relation, identification column of any one table can be added in to another table. This column is used in the join operation.

One- to- many: In this type of relation, identification column of the table of 'one' side of the relation is added in to the table on 'many' side of relation. For example to join table of basin (one) to dam (many), the identification column of basin is placed in the dam table.

Many- to- many: In this type of relationship, an intermediate table is prepared. This table contains, identification column of both the tables, which have this relationship type. There after all three tables (to tables to be joined and the intermediate table) are joined using these identification numbers.

Standard Query language (SQL)

SQL is a scripting language used for defining and processing data bases. These may be executed as commands within DBMS. In DBMS software, visual interfaces are also provided, which are equivalent to SQL commands. The SQL can also be used in scripting languages, e.g. VBScript etc., or languages, e.g. java, C# etc. SQL was developed in 1970. ANSI adopted the SQL as standard in 1992. This standard is called SQL- 92.

1.8 APPLICATIONS

<u>Water Resources Information System (WRIS)</u>

A WRIS (USA) is hosted by USGS and US EPA. Tabular information, maps, hydrometeorological data of watershed can be extracted from the system. A web based application is developed by CWC for hydrometric information dissemination. CWC monitors discharge in various rivers and availability of gauge/ discharge information can be known from the system. River basin map, monitoring location etc. are displayed in map form.

<u>Hydrological modeling and GIS</u>

A hydrological model is a mathematical model used to simulate river or stream flow and calculate water quality parameters. Runoff estimation is useful in water resources planning, design of hydraulic structures, bridges etc. Impact of land use/ climatic changes on hydrological cycle can be studied through simulation studies. Inundation mapping can also be done in river basin using hydrologic/ hydraulic mathematical models.

Remotely sensing data are important data source in modeling. Land use and its management practices affect runoff from basins. Land use and cover maps, up to level IV, can be prepared using remotely sensed data. Soil texture and types can also be interpreted. SRTM (Shuttle Radar Topographic Mapping Mission) Digital elevation models (DEM) are available at 90 m spatial resolution and 1 m least count for the world. Rainfall data are available at 15' geographic and 3 hourly temporal resolution from TRMM (Tropical Rainfall Measuring Mission). Water and land temperature can be monitored using satellite data.

GIS is a useful tool in data preparation for hydrologic models. DEM can be used for extracting catchment and channel characteristics in GIS. Semi- distributed models require division of basin in to sub basins. Sub basins can be delineated automatically using DEM.

Catchment characteristics, namely, area, perimeter, longest flow path, flow path up to the centroid, average slope can be computed using DEM in GIS. For channels, length, average slope and cross section can be extracted. In GIS framework, set up files for the models can be prepared. These files can be used in model simulation. Examples of ARC View GIS interfaces are GeoHMS, GeoRAS, AvSWAT, Mike 11 GIS for HEC HMS, HEC RAS, SWAT and Mike 11 hydrological models respectively (Table 1.4).

Table 1.4 Hydrological Models

Model	GIS	Remark
TRIWACO (Quasi 3-D FEM GW model)	ESRI	FEM equilateral triangles were generated
WetSpa (Physically based distributed catchment model)	Embedded	Hydrological processing of DEM is done
Travel time method	ArcView	Hydrological processing of DEM is done
HSPF/ SWAT	Basin 3.0, AvSWAT, ArcSWAT	-
TOPMODEL/ WMS 5.1	Computer programs	GRIDATB (for Topographic Index) TOPAZ (For hydrological processing of DEM)
Modflow/ MT3D	ArcView (with 3D Analyst)	Generate mesh, pre, post processing, Launch of model
HYDROTEL (Distributed/ semi-distributed catchment model)	Embedded	PHYSITEL- Hydrological processing of DEM and computation of internal drainage structure) Interpolation
HEC- HMS	CRWR- PrePro GeoHMS ArcView, Spatial Analysis	DEM processing, topologic structure generation (junction, sink, sub basin and reach), parameter generation, export to ASCII file (basin model)

Arc View is widely used software for providing interface for data preparation for hydrological models. Extensions are written for several hydrological models. These extension are written using scripting language namely Arc Macro Language (AML). To use an extension, the extension is required to be activated in Arc View menu. Special menu items and icons are added after the extension is selected. The menu items/ icons can be used for data preparation of the hydrological models.

Groundwater studies

Groundwater depth and quality is studied in GIS through kriging interpolation technique. Water quality variables e.g. EC, RSC, HCO3 etc. can be interpolated using the technique. In these applications, sample locations can be important. For example, samples are taken from working tube wells, which may be in general of good quality and poor water-quality is under represented. Range can be of order of 10 to 100 km. Models can be exponential, spherical etc.

Groundwater potential and quality can be mapped in GIS environment. Various layers namely slope, geology, hydromorphogeology, distances to drainage channel, tanks and lineaments, depth to water table, depth of weathered zone can be overlaid and integrated on GIS environment to obtain groundwater potential map. Similarly, layers namely magnesium, incrustation problem, TH, TDS can be integrated to obtain quality map in incrustation and corrosion problem areas.

Reservoir sedimentation

A reservoir changes the hydraulics of flow by forcing the energy gradient to approach zero. This results in a loss of silt transport capacity of the water. Larger particles are deposited first. The smaller particles move longer distances in to the reservoir before depositing. There is a higher deposition rate compared to pre dam river scenario. With passage of time, the useful storage capacity of the reservoirs reduces, resulting in reduced benefits. A delta formation occurs in the reservoirs as depicted in Fig. 5.

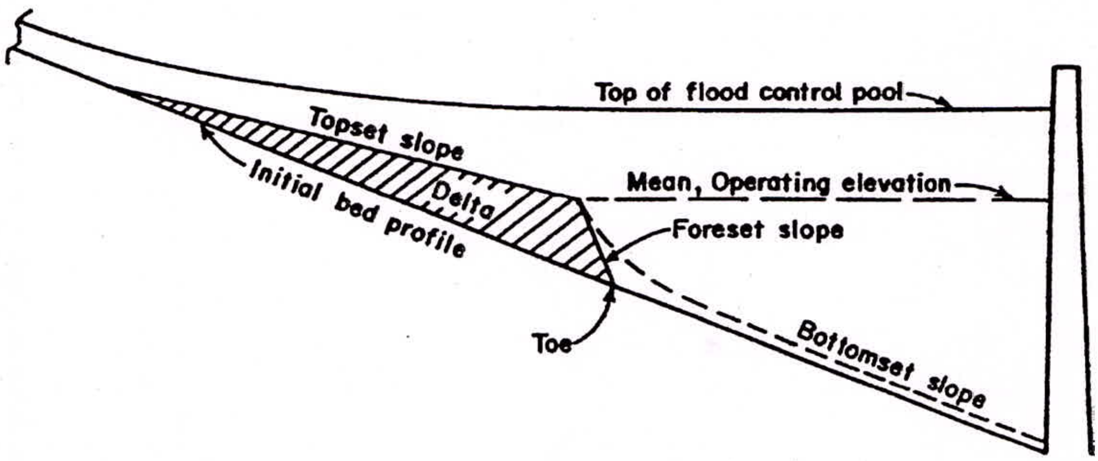

Fig. 5 Sediment deposition in the reservoirs

As per the recommendation Reservoir Sedimentation Committee of Ministry of Agriculture and Irrigation (Department of Irrigation), Government of India (1982), capacity surveys on regular intervals of once in 5 years (Mahto and Patil 2005). In practice, surveys are done at an interval of 2 to 15 years (Shangle 1991). Remote sensing can be crucial role in cost effective sedimentation surveys of reservoir's live storages. The techniques relies on estimation of water spread area from digital analysis of satellite data and collecting corresponding reservoir elevation data recorded normally at a daily interval by project authorities. Multidate satellite data are used to populate revised elevation- area table. The table is expanded by interpolating the area corresponding to 0.1 m elevation interval. Cone formula is used to estimate incremental volume. From the incremental volume the cumulative reservoir capacity at various elevations is computed.

$$V = \frac{h}{3}\left(a_1 + a_2 + \sqrt{a_1 \times a_2}\right)$$

Where, V= incremental volume, a1 and a2 = water spread area at two neighbouring elevations, h= elevation interval

Land use classification

Information on existing land use and their changes over time is needed in water resources planning and management, amelioration of land degradation etc. Information is available from past land use maps, topographic maps etc. Due to time variant nature of this geographical data, satellite remote sensing is very useful.

Land use refers to, man's activities on land which are directly related to the land. Land cover, on the other hand, describe, the vegetation and artificial constructions covering the land surface. USGS (United States Geological Survey) has developed a land use and land cover classification system for use with remotely sensed data (Table 1.5). The system has four hierarchical levels, namely, I to IV. Levels I and II (Table 2) are useful for national, statewide classification. Other two levels are useful for intrastate, regional, county and municipal level classification. Latter levels classes are developed by users. First two levels have standard and well defined classes. Level I and II classes can be delineated using medium resolution satellite data, e.g. IRS LISS- III, Landsat TM etc. Level III and IV classes can be delineated using fine resolution data, e.g. IRS LISS- IV, Panchromatic, aerial photographs etc.

Multi date satellite remotely sensed data in single season may be used in crop type discrimination. In particular, satellite data may be very useful in discrimination in paddy (seasonal) and sugarcane (perennial) crops. Staggered sowing and different duration in crops pose problem is accurate classification of crop types. Horticulture areas have different signature than crops.

Table 1.5 USGS land use and land cover classification system

Level I	**Level II**
1 Urban or Built-up Land	11 Residential 12 Commercial and Services 13 Industrial 14 Transportation, Communications, and Utilities 15 Industrial and Commercial Complexes 16 Mixed Urban or Built-up Land 17 Other Urban or Built-up Land
2 Agricultural Land	21 Cropland and Pasture 22 Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas 23 Confined Feeding Operations 24 Other Agricultural Land
3 Rangeland	31 Herbaceous Rangeland 32 Shrub and Brush Rangeland 33 Mixed Rangeland
4 Forest Land	41 Deciduous Forest Land 42 Evergreen Forest Land 43 Mixed Forest Land
5 Water	51 Streams and Canals 52 Lakes 53 Reservoirs 54 Bays and Estuaries
6 Wetland	61 Forested Wetland 62 Nonforested Wetland
7 Barren Land	71 Dry Salt Flats.

	72 Beaches 73 Sandy Areas other than Beaches 74 Bare Exposed Rock 75 Strip Mines Quarries, and Gravel Pits 76 Transitional Areas 77 Mixed Barren Land
8 Tundra	81 Shrub and Brush Tundra 82 Herbaceous Tundra 83 Bare Ground Tundra 84 Wet Tundra 85 Mixed Tundra
9 Perennial Snow or Ice	91 Perennial Snowfields 92 Glaciers

Land degradation

Irrigated agriculture areas often face problems of water logging and salinity. The problems are refereed as twin problem as waterlogging leads to soil salinization in long run. Identification of extent of the affected area is pre requisite for reclamation. Criteria adopted for water logging are given in Table 1.6. Criteria for salinity and sodicity are listed in Table 1.7.

Area with surface pondage and moist soil can be delineated easily using remote sensing data. Water has black tone in standard FCC in visible and near IR bands. Most soil has dark signature in these imageries. Shallow water table conditions often are not detected using optical remotely sensed data unless its expression is visible on the surface of the earth. Areas where yield is affected can be monitored.

Table 1.6 Criteria adopted for waterlogging

Waterlogging	National Commission on Agriculture (1976)	Ministry of Water Resources, GoI (1991)
Waterlogged/ Critical	Water table < 1.5 m	Water table < 2 m
Potentially waterlogged		Water table 2-3 m
Safe area		Water table > 3 m

Table 1.7 Criteria for soil salinity/ sodicity

Degree of salinity/ sodicity	Salinity EC (dSm^{-1})	Sodicity	
		pH	ESP
Slight	4-8	8.2-9.0	< 15
Moderate	8-25	9.0-9.8	15-40
Strong	>25	>9.8	>40

Saline areas possess salt efflorescence on the surface. Due to this, saline areas have bright appearance in optical remote sensing. Sodic area have different signature than saline areas. In sodic areas, the infiltration is very less and thus water gets stagnant in the areas and thus the area can be identified through surface pondage and moist soil. Both, saline and sodic areas have poor growth of vegetation. Waterlogged areas can also be delineated using GIS technique using water depth map. The map is processed to correct any discrepancies in depths (e.g. negative values). The maps can be utilized to classify areas as waterlogged/ crtitical, potential waterlogged and safe.

Table 1.8 Runoff Curve Numbers for Urban Areas

Cover Type and Hydrologic Condition	Average Percent Impervious Area	A	B	C	D
Open space (lawns, parks, golf courses, cemeteries, etc.)					
• Poor condition (grass cover < 50%)		68	79	86	89
• Fair condition (grass cover 50% to 75%)		49	69	79	84
• Good condition (grass cover > 75%)		39	61	74	80
Paved parking lots, roofs, driveways, etc. (excluding right-of-way)		98	98	98	98
Streets and roads:					
• Paved; curbs and storm drains (excluding right-of-way)		98	98	98	98
• Paved; open ditches (including right-of-way)		83	89	92	93
• Gravel (including right-of-way)		76	85	89	91
• Dirt (including right-of-way)		72	82	87	89
Urban districts:					
• Commercial and business	85	89	92	94	95
• Industrial	72	81	88	91	93
Residential districts by average lot size:					
• 1/8 acre or less (town houses)	65	77	85	90	92
• 1/4 acre	38	61	75	83	87
• 1/3 acre	30	57	72	81	86
• 1/2 acre	25	54	70	80	85
• 1 acre	20	51	68	79	84
• 2 acres	12	46	65	77	82
Developing urban areas:					
Newly graded areas (pervious areas only, no vegetation)		77	86	91	94

Table 1.9 Runoff Curve Numbers for Cultivated Agricultural Land						
Cover Type	**Treatment**	**Hydrologic Condition**	**A**	**B**	**C**	**D**
Fallow	Bare soil		77	86	91	94
	Crop residue	Poor	76	85	90	93
	cover (CR)	Good	74	83	88	90
Row Crops	Straight row (SR)	Poor	72	81	88	91
		Good	67	78	85	89
	SR + CR	Poor	71	80	87	90
		Good	64	75	82	85
	Contoured (C)	Poor	70	79	84	88
		Good	65	75	82	86
	C + CR	Poor	69	78	83	87
		Good	64	74	81	85
	Contoured & terraced (C&T)	Poor	66	74	80	82
		Good	62	71	78	81
	C&T + CR	Poor	65	73	79	81
		Good	61	70	77	80
Small grain	SR	Poor	65	76	84	88
		Good	63	75	83	87
	SR + CR	Poor	64	75	83	86
		Good	60	72	80	84
	C	Poor	63	74	82	85
		Good	61	73	81	84
	C + CR	Poor	62	73	81	84
		Good	60	72	80	83
	C&T	Poor	61	72	79	82
		Good	59	70	78	81
	C&T + CR	Poor	60	71	78	81

		Good	58	69	77	80
Close-seeded	SR	Poor	66	77	85	89
or broadcast		Good	58	72	81	85
Legumes or C		Poor	64	75	83	85
Rotation		Good	55	69	78	83
Meadow	C&T	Poor	63	73	80	83
		Good	51	67	76	80

Table 1.10 Runoff Curve Numbers for Other Agricultural Lands

Cover Type	Hydrologic Condition	A	B	C	D
Pasture, grassland, or range-continuous forage for grazing	Poor Fair Good	68 49 39	79 69 61	86 79 74	89 84 80
Meadow – continuous grass, protected from grazing and generally mowed for hay		30	58	71	78
Brush – brush-weed-grass mixture, with brush the major element	Poor Fair Good	48 35 30	67 56 48	77 70 65	83 77 73
Woods – grass combination (orchard or tree farm)	Poor Fair Good	57 43 32	73 65 58	82 76 72	86 82 79
Woods	Poor Fair Good	45 36 30	66 60 55	77 73 70	83 79 77
Farmsteads – buildings, lanes, driveways, and surrounding lots		59	74	82	86

Pasture: Poor is < 50% ground cover or heavily grazed with no mulch, Fair is 50% to 75% ground cover and not heavily grazed, and Good is >75% ground cover and lightly or only occasionally grazed.

Meadow: Poor is <50% ground cover, Fair is 50% to 75% ground cover, Good is >75% ground cover.

Woods/grass: RCNs shown were computed for areas with 50 percent grass (pasture) cover. Other combinations of conditions may be computed from RCNs for woods and pasture.

Woods: Poor is forest litter, small trees, and brush destroyed by heavy grazing or regular burning. Fair is woods grazed but not burned and with some forest litter covering the soil. Good is woods protected from grazing and with litter and brush adequately covering soil.

Table 1.11 Rainfall Groups for Antecedent Soil Moisture Conditions during Growing and Dormant Seasons

Antecedent Condition	Description	Growing Season 5-Day Antecedent Rainfall	Dormant Season 5-Day Antecedent Rainfall
Dry AMC I	An optimum condition of watershed soils, where soils are dry but not to the wilting point, and when satisfactory plowing or cultivation takes pace	Less than 1.4 in. or 35 mm	Less than 0.05 in. or 12 mm
Average AMC II	The average case for annual floods	1.4 in. to 2 in. or 35 to 53 mm	0.5 to 1 in. or 12 to 28 mm
Wet AMC III	When a heavy rainfall, or light rainfall and low temperatures, have occurred during the five days previous to a given storm	Over 2 in. or 53mm	Over 1 in. or 28 mm