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

GROUNDWATER STUDIES USING GIS

By

VIJAY KUMAR
NIH, ROORKEE

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

GROUNDWATER STUDIES USING GIS

1.0 INTRODUCTION

Groundwater is one of the most valuable natural resources, which supports human health, economic development and ecological diversity. Because of its several inherent qualities (e.g., consistent temperature, widespread and continuous availability, limited vulnerability, low development cost, drought reliability, etc.), it has become an immensely important and dependable source of water supplies in all climatic regions including both urban and rural areas of developed and developing countries. The distribution of water on the earth is highly unbalanced. Nearly 97.41% of water is confined to the world's oceans and are unsuitable for human and livestock consumption. Major portion of the remaining 2.59% is locked up in the glaciers (1.953%) and beneath the surface as groundwater (0.614%), thus leaving only a meager 0.015% in the rivers and lakes for consumption by terrestrial inhabitants. This quantity of water was enough to support civilization. But, in recent years due to population explosion and urbanization, water resources from rivers, ponds and lakes have become inadequate. Therefore, an urgent need has arisen for exploring and managing surface as well as groundwater resources for continuous and dependable water supply for the growing needs of population. It is now well-recognized fact that water is finite and vulnerable resource, and it must be used efficiently and in an ecologically sound manner for present and future generations.

The geographic information system (GIS) has emerged as an effective tool for handling spatial data and decision making in several areas including engineering and environmental fields. Remotely sensed data are one of the main sources for providing information on land and water related subjects. These data being digital in nature, can be efficiently interpreted and analyzed using various kinds of software packages. It is easy to feed such information into a GIS environment for integration with other types of data and then do analyses. The combined use of remote sensing and GIS is a valuable tool for the analysis of voluminous hydrogeologic data and for the simulation modeling of complex subsurface flow and transport processes under saturated and unsaturated conditions. Undoubtedly, the GIS technology allows for swift organization, quantification and interpretation of a large volume of hydrologic and hydrogeologic data with computer accuracy and minimal risk of human errors.

It is widely recognized that GIS provides a large range of analytical capabilities to operate on topological relationships or spatial aspects of the geographical data, on the non-spatial attributes of such data, or on non-spatial and spatial attributes combined. GIS facilitates the integration of disparate data sets, creation of new and derivative data sets, and development and analysis of spatially explicit variables. Furthermore, the integration of GIS with spatial statistical analysis has the potential to become a powerful analytical toolbox, enabling regional and social scientists to gain fundamental insight into the nature of spatial structures of regional development (Brown, 1996).

A linkage between GIS and spatial data analysis is considered to be an important aspect in the development of GIS into a research tool to explore and analyze spatial relationships. Simply put, the power of a GIS as an aid in spatial data analysis lies in its geo-relational data base structure, i.e., in the combination of value information and location information. The link between these two allows for the fast computation of various characteristics of the spatial arrangement of the data, such as the contiguity structure between observations, which are essential inputs into spatial data analysis. The GIS also provides a flexible means to "create new data," i .e., to transform data between different spatial scales of observation, and to carry out aggregation, partitioning, interpolation, overlay and buffering operations. Of course, such "data" is nothing but the result of computations, themselves based on particular algorithms that often use parameter estimates and model calibrations obtained by statistical means. The powerful display capabilities contained in a GIS also provide excellent tools for the visualization of the results of statistical analyses.

In the context of groundwater studies, GIS technology is considered useful as it facilitates handling of diverse type of spatial information, e.g. topographic maps, landform maps, geological maps, various contour maps of water table and water quality etc. and offers flexibility of operation, speedy proceeding and higher accuracy.

Preliminary work in ground water modeling requires the translation and transform of information on maps, charts, and tables into computer readable form. The work is lengthy, tedious, and error prone. Changes required in the data sets during the model calibration often involves sifting through thousands of numbers (data) to make what often turned out to be minor modifications to the input data sets.

The specification of hydrological information such as rainfall, parameter information such as hydraulic conductivity, design parameter specification such as well locations and discharge values, and auxiliary conditions such as boundary conditions all involve the organization and manipulation of enormous quantities of data. Virtually all of this information is spatially, and in some instances temporally distributed. Much of it is available in computerized database either as maps in raster or vector format or as data tables. Due to advantages in computer-graphical technology, the information in such database is now accessed most efficiently through GIS systems.

Collection of large volumes of geographical data required for ground water modeling is very laborious if done by hand. For both the pre processing as well as the post processing stage, the use of the GIS saves much time and becomes possible to improve more results. In general the input parameters for existing hydrologic models are prepared in the GIS and passed on to the model via an interface.

2.0 SPATIAL INTERPOLATION USING GIS

The sustainable management of groundwater resources needs quantitative information on its behaviour in space and time. As groundwater use has increased, issues associated with the quality of groundwater resources have likewise grown in importance. The groundwater data sets typically contain many variables measured at several spatially scattered locations. Knowledge of spatial variability of groundwater data is essential for making reliable groundwater interpretations and for making accurate predictions of groundwater data at any particular location in the aquifer. With the availability of distributed hydrological models, which can handle large volume of data, the spatial information of hydrological data at a grid pattern is not only useful but is necessary to get reliable results. Various methods, both simple and complicated, are available and are in use to get the information about any parameter on a pre-specified grid when the parameter is measured at random points in the field. Spatial interpolation is the process of using points with known values to estimate values at other points. GIS can provide effective spatial analysis capabilities required to use various data in modeling studies. The various methods of interpolation are discussed below.

2.1 Regression Model

A global interpolation method, regression model relates a dependent variable to a number of independent variables in a linear equation (an interpolator), which can then be used for prediction or estimation. The trend surface analysis, a type of regression model, approximates points with known values with a polynomial equation. The equation or the interpolator can then be used to estimate values at other points. A linear or first-order trend surface uses the equation:

$$z_{xy} = b_0 + b_1 x + b_2 y \qquad (1)$$

where, the attribute value z is a function of x and y coordinates. The b coefficients are estimated from the known points. Higher-order trend surface models are required to approximate more complex surfaces. A cubic or a third-order model, for example, includes hills and valleys. A cubic trend surface is based on the equation:

$$z_{xy} = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 y^2 + b_5 xy + b_6 x^3 + b_7 y^3 + b_8 x^2 y + b_9 xy^2 \qquad (2)$$

A third-order trend surface required estimation of 10 coefficients (i.e., b_i), compared to three coefficients for a first-order and six coefficients for second-order surface. A higher order trend surface model therefore required more computation than a lower-order model does.

In GIS (ILWIS software) this method is known as trend surface method. This method calculates pixel values by fitting one surface through all point values in the map. The surface may be of the first order up to the sixth order. A trend surface may give a general impression of the data. Surface fitting is performed by a least squares fit.

2.2 Thiessen Polygon

Thiessen polygons assume that any point within a polygon is closer to the polygon's known point than any other known points. Thiessen polygons do not use an interpolator but require initial triangulation for connecting known points. Each known point is connected to its nearest neighbors. Thiessen polygons are easily constructed by connecting lines drawn perpendicular to the sides of each triangle at their midpoints.

In GIS (ILWIS software) this method is known as nearest point method. This method assigns to pixels the value, identifier or class name of the nearest point, according to Euclidean distance. This method is also called Nearest Neighbour.

2.3 Inverse Distance Weighted Interpolation

Inverse distance weighted (IDW) interpolation is an exact method that enforces that the estimated value of a point is influenced more by nearby known points than those farther away. The general equation for the IDW method is:

$$Z_0 = \frac{\sum_{i=1}^{n} Z_i \frac{1}{d_i^k}}{\sum_{i=1}^{n} \frac{1}{d_i^k}} \qquad (3)$$

where Z_0 is the estimated value at point 0, Z_i is the Z value at known point *i*, d_i is the distance between point *i* and point 0, n is the number of known points used in estimation, and *k* is the specified power. The power *k* controls the degree of local influence. A power of 1.0 means a constant rate of change in value between points (linear interpolation). A power of 2.0 or higher suggests that the rate of change in values is higher near a known point and levels off away from it. The degree of local influence also depends on the number of known points used in estimation. An important characteristic of IDW interpolation is that all predicted values are within the range of maximum and minimum values of the known points.

In GIS (ILWIS software) this method is known as Moving Average method. The Moving average operation is a point interpolation which requires a point map as input and returns a raster map as output. The values for the output pixels are the weighted averages of input point values. Weighted averaging is the calculation of the sum of the products of weights and point values, divided by the sum of weights. The weight factors for the input points are calculated by a user-specified weight function. This method ensure that points which are close to an output pixel obtain large weights and that points which are farther away from an output pixel obtain small weights. Values of points which are close to an output pixel are thus of greater importance to this output pixel value than the values of points which are farther away. By specifying a limiting distance, you can influence until which distance from any output pixel, points will be taken into account for the calculation a value for that output pixel; for each output pixel, only the values of the points which fall within the limiting distance to this output pixel will be used. Values of points that are farther away from an output pixel than the specified limiting distance, obtain weight zero by the weight calculation, and these values will thus not be used in the output pixel value calculation.

2.4 Kriging

Kriging is an alternative to many other point interpolation techniques. Unlike straightforward methods discussed above, Kriging is based on a statistical method. Kriging is the only interpolation method available in ILWIS that gives you an interpolated map and output error map with the standard errors of the estimates. Kriging is such a technique, which takes into consideration the spatial structure of the parameter and so scores over the other methods. Kriging is based on the Theory of Regionalized Variables. When a variable is distributed in space, it is said to be "regionalized". All the parameters generally used in groundwater hydrology, such as transmissivity, hydraulic conductivity, piezometric heads, vertical recharge etc. can be called regionalized variables.

Semivariogram

Kriging uses the semivariogram to measure the spatial correlated components, a component that is also called spatial dependence or spatial autocorrelation. The semivariogram is half of the arithmetic mean of the squared difference between two experimental measures, ($Z(x_i)$ and $Z(x_i+h)$), at any two points separated by the vector h. Before values of any parameter can be estimated with kriging, it is necessary to identify the spatial correlation structure from the semivariogram, which shows the relationship between semivariance and the distance between sample pairs.

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \qquad (4)$$

where $\gamma^*(h)$ = estimated value of the semivariance for lag h; N(h) is the number of experimental pairs separated by vector h; $z(x_i)$ and $z(x_i+h)$ = values of variable z at x_i and x_i+h, respectively; x_i and x_i+h = position in two dimensions. A plot of $\gamma^*(h)$ versus the corresponding value of h, also called the semivariogram, is thus a function of the vector h, and may depend on both the magnitude and the direction of h. A sample plot of semivariogram is shown in Fig. 1.

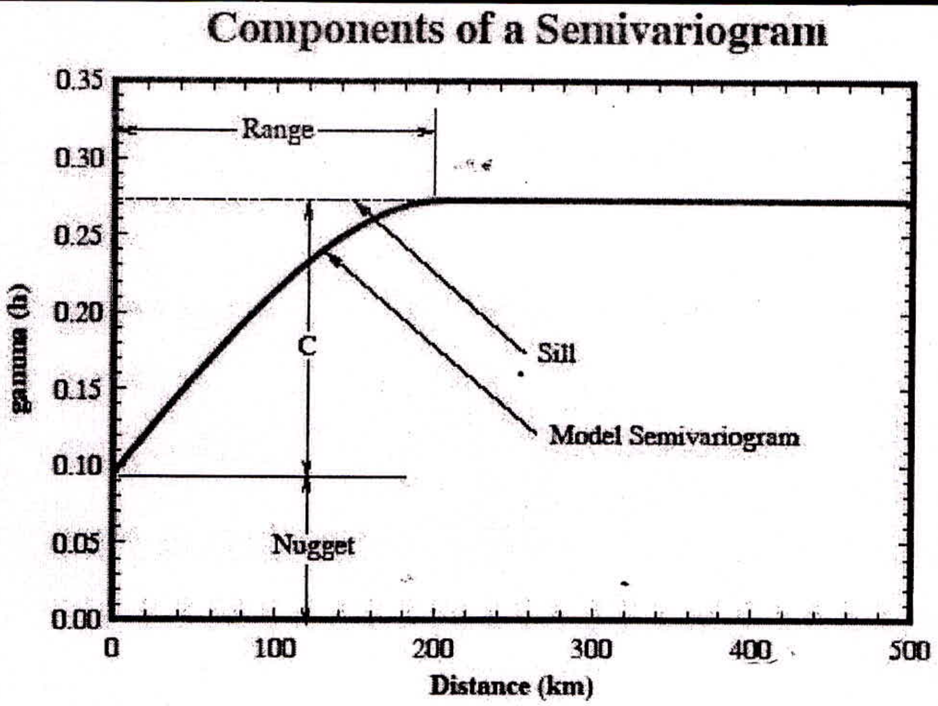


Fig. 1 Sample Plot of Semivariogram

The distance at which the variogram becomes constant is called the range, *a*. It is considered that any data value Z(x) will be correlated with any other value falling within a radius, *a* and thus range corresponds to the zone of influence of the RV. The value of the semivariogram at a distance equal to the range is called the sill. Semivariograms may also increase continuously without showing a definite range and sill. The value of the semivariogram at extremely small separation distance is called the nugget effect.

Structural Analysis

The observed data is used to calculate the experimental semivariogram. A mathematical function used to approximately represent this semivariogram is known as the theoretical semivariogram. Some of the theoretical semivariogram models are (Fig. 2).

Spherical model: -

$$\gamma(h)=\begin{cases} C_0[1-\delta(h)]+C\left[\frac{3}{2}\frac{h}{a}-\frac{1}{2}\frac{h^3}{a^3}\right] & h\le a \\ C_0+C & h>a \end{cases} \tag{5}$$

Exponential model: -

$$\gamma(h)=C_0[1-\delta(h)]+C\left[1-\exp\left(-\frac{h}{a}\right)\right] \tag{6}$$

Gaussian model: -

$$\gamma(h)=C_0[1-\delta(h)]+C\left[1-\exp\left(-\frac{h^2}{a^2}\right)\right] \tag{7}$$

Linear model: -

$$\gamma(h)=C_0[1-\delta(h)]+bh \tag{8}$$

where, δ(h) is the Kronecker delta = $\begin{cases}1 & h=0\\ 0 & h\neq 0\end{cases}$, C_0 is the Nugget effect, C_0+C is the Sill, *a* is the Range and *b* is the slope.

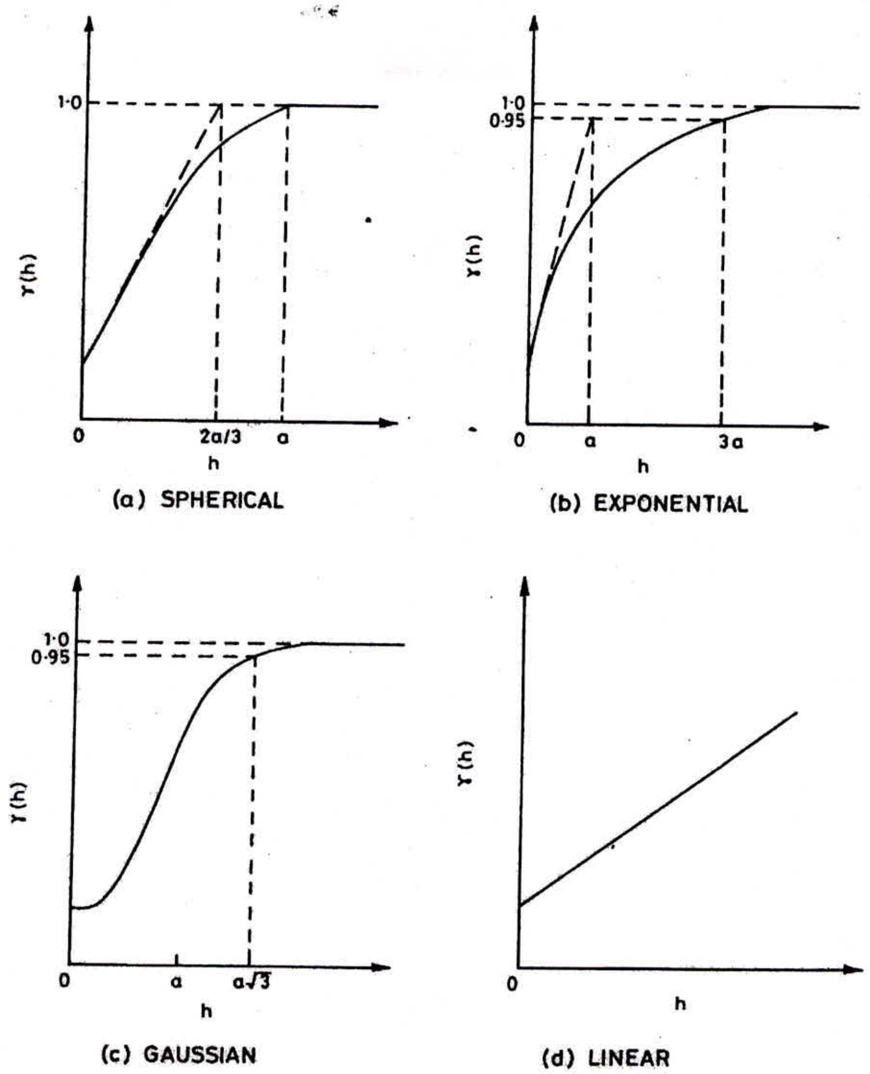


Fig. 2 Theoretical models of semivariogram

Kriging

Kriging is a geostatistical method for spatial interpolation. It is a technique of making optimal, unbiased estimates of regionalized variables at unsampled locations using the structural properties of the semivariogram and the initial set of data values. Consider a situation in which a property is measured at a number of points, x_i , within a region to give values of $z(x_i)$, i=1,2,3,...,N. (x_i is the coordinate of the observation point in 1, 2 or 3-dimensional space). From these observations, the value of the property at any place x_0 can be estimated as

$$z^*(x_0) = \sum_{i=1}^{N} \lambda_i z(x_i) \qquad i=1,2,3,\ldots\ldots,N \tag{9}$$

where,

$z^*(x_0)$ = estimated value at x_0
λ_i = weights chosen so as to satisfy suitable statistical conditions
$z(x_i)$ = observed values at points x_i, i= 1,2,3,.......,N
N = sample size

In kriging, the weights λ_i are calculated so that $Z^*(x_0)$ is unbiased and optimal i.e.

$$E\{Z^*(x_0) - Z(x_0)\} = 0 \tag{10}$$

$$Var\{Z^*(x_0) - Z(x_0)\} = \text{minimum} \tag{11}$$

The best linear unbiased estimate of $Z(x_o)$ is obtained by using Lagrangian techniques to minimise Eq. (11) and then optimising the solution of the resulting system of equations when constrained by the nonbiased condition of Eq. (10). The following system of equations, known as kriging system, results from the optimization:

$$\begin{cases} \sum_{j=1}^{N} \lambda_j \gamma(x_i, x_j) + \mu = \gamma(x_i, x_0) & i = 1,2,3,\ldots\ldots, N \\ \sum_{j=1}^{N} \lambda_j = 1 \end{cases} \tag{12}$$

where,

μ = Lagrange multiplier

$\gamma(x_i, x_j)$ = semivariogram between two points x_i and x_j

Solution of the above set provides the values of λ_i, which can be used with Eq. 9 for estimation. The minimum estimation variance, or kriging variance, is written as:

$$\sigma_k^2(x_0) = \sum_{i=1}^{N} \lambda_i \gamma(x_i, x_0) + \mu \tag{13}$$

3.0 GIS IN GROUNDWATER MODELLING

Ground-water simulation models are important tools in water-resources planning and management. For regional groundwater flow studies, numerical models (such as finite-difference or finite-element models) are often used to represent heterogeneous hydrogeological parameters and boundary conditions. These numerical models, however, often require large data sets, which are difficult to manage. Assembling data, assigning parameters for each model cell, and checking for errors are time-consuming tasks. As a result, once a model grid or mesh is designed and parameters assigned, there is generally no consideration of model sensitivity to discretization. Moreover, the number of simulations performed during model calibration and verification (testing the model with an independent set of data) is often limited. Finally, the visualization and translation of model results can be a tedious and time-consuming task.

Geographic Information Systems (GIS's) are technologies that can greatly facilitate the development, calibration, and verification of ground-water models, as well as the display of ground-water model parameters and results. Through the linking of a digital mapping system to a database, GIS's have the ability to integrate data layers and perform spatial operations on data. Thus, GIS's can automate many of the data compilation and management duties in ground-water modeling. Spatial statistics and grid design capabilities of GIS's can improve the modeling effort and aid in reliability assessment. Furthermore, GIS's have many visual display capabilities that can aid in calibration, verification, and the production of final results.

Some common uses of GIS in groundwater modeling have been as follows:

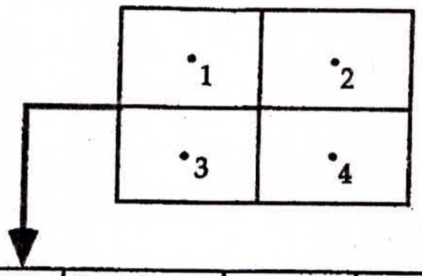
1. Preparation of data for model input. Contours or triangular irregular networks (TINs) created from point-data coverages of aquifer properties are intersected with the model grid. Effective parameters for each cell are assigned automatically and systematically.
2. Assessment of the adequacy of the model input through the visual display andlor comparison of contour or TIN values (such as in making sure the top of the aquifer is not above the land surface).
3. Allocation of pumping and recharge rates to each grid cell.
4. Visual comparison of simulated and measured head or concentration values.
5. Interactive revision of parameter values and/or spatial discretization.
6. Display of model results such as contours of hydraulic head, flow vectors, and water quality contours.

Two common methods of ground-water modeling are the finite-difference method and the finite-element method. In two-dimensions the finite-difference method employs a rectangular discretization in which parameter values are entered and hydraulic head is computed for each cell. For example, Fig 3 shows how the extent of an aquifer is represented in the ground-water model MODFLOW. In a vector-based GIS, geographic features are represented a points, lines, and areas, and the locations of the features are referenced by a Cartesian *(x, y)* coordinate system. Thus, each point corresponds to a single *x, y* location, each line is defined by a series of *x, y* locations, and each area is defined by a set of line segments enclosing it. In a raster-based GIS, reality is divided into a rectangular grid organized as a set of rows and columns. Each cell of the grid has a value representing a geographic phenomenon (such as elevation, hydraulic conductivity) and the cells are referenced by their row and column. Thus, a Vector GIS can represent feature shape more accurately, but raster based GIS can better represent gradual transition between features and surfaces. Both raster- and vector-based GIS can represent the finite-difference grids. In fact, a raster-based GIS is directly analogous to a finite-difference grid with a GIS data layer representing each model parameter, that is, one grid may represent the spatial distribution of hydraulic conductivity, another the elevation of the base of the aquifer, and so on. A vector-based GIS can represent a finite-difference grid as a point of coverage by locating a point in the center of each finite-difference cell and storing for each point a database record that contains all the model parameters attributed to that cell (Fig. 4).

0	0	0	0	0	0	0	0	0
0	0	0	-1	-1	-1	-1	0	0
0	0	-1	1	1	1	1	1	0
0	-1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	0	0
0	1	1	1	-1	-1	1	1	0
0	1	1	1	-1	1	1	0	0
0	0	1	1	1	1	1	0	0
0	0	1	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0

0: no-flow boundary
1: water table cell
-1: constant-head boundary

FIG. 3. Plan View of MODFLOW Aquifer Representation



ID	Hyd. Cond.	Base	Top	Type
1	0.0	999.9	999.9	0
2	68.3	416.7	455.9	1
3	85.0	422.0	460.1	-1
4	91.2	418.8	457.5	1

FIG. 4. Point Coverage and Attribute Table for Finite-Difference Model

In finite-element ground-water models, triangular elements are often used to represent two-dimensional space. Parameter values are entered for each element, and hydraulic head is determined at each node, therefore allowing the spatial interpolation of head along the surface of each element. Though more complex than the finite-difference method, the finite-element method can better represent curved boundaries. As shown in Fig. 5, a vector-based GIS can represent a finite-element mesh as a polygon coverage and a point coverage. The database record for each point and polygon contains the model parameters for each node and element, respectively, and the GIS maintains the topological relationship between each node and its corresponding elements.

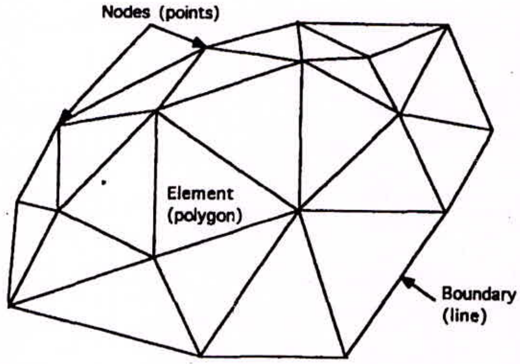


Fig. 5: Two-Dimensional Finite-Element Aquifer Representation

During the calibration of groundwater model, GIS can play a very effective and crucial role. Fig. 6 shows how parameter assignment is automated with GIS during groundwater modelling.. Given a contour map in the form of a GIS line coverage, a TIN (shown in Fig. 4 as a 3D surface) can be created and intersected with the point coverage that represents the model grid. The TIN value at the location of each point (model grid cell) is then automatically assigned to the proper item in the point attribute table. For instance, given a digitally contoured map of hydraulic conductivity, a TIN can be created and intersected with the model grid point coverage to assign automatically a conductivity value to each model cell. If the model grid is resized, the TIN is simply intersected with a new point coverage and the parameter values automatically reassigned. Alternatively, the average hydraulic conductivity value for each grid cell can be determined by calculating a weighted average of values from the TIN over each cell.

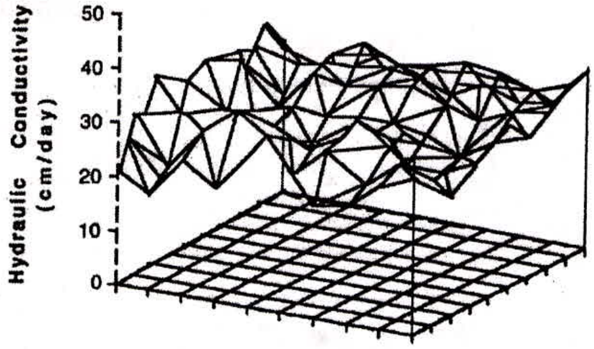


Fig. 7 shows the usefulness of GIS in calibrating a regional ground-water flow model. Contours of observed and simulated head can be overlaid for quick visual comparison, or else each contour plot can be converted to a TIN for statistical analysis. GIS can calculate the volumetric differences between the TINs, the maximum and minimum differences, and the difference at a particular point of interest. Also, the differences in slope along particular lines can be computed.

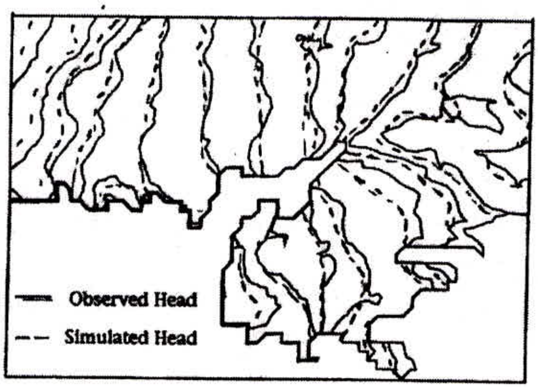


Fig. 7. Overlaying observed and simulated heads for model

GIS can provide a powerful platform for displaying model results. The model results can be viewed by displaying colour maps or contour maps of output elements like groundwater heads, water table, drawdown and species concentration etc.

** *** **